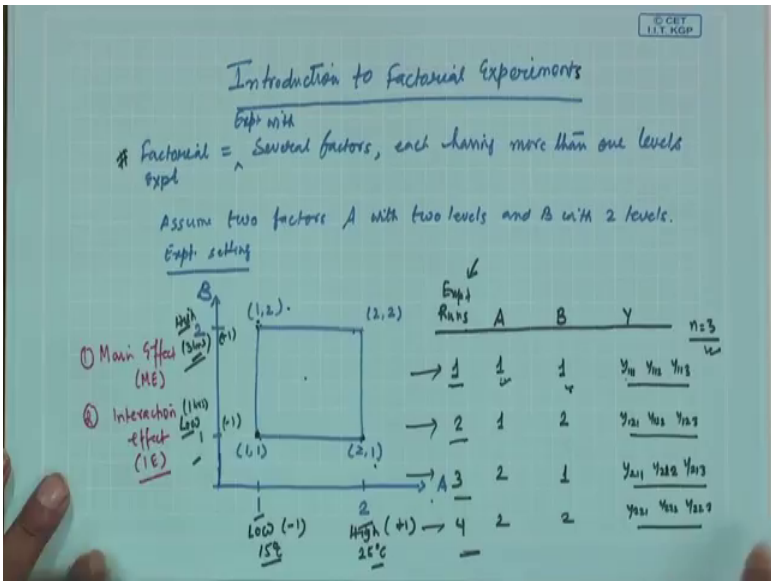
Design and Analysis of Experiments

Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 30
Introduction to Factorial Experiments

Welcome. Today will discuss factorial experiments, factorial experiments.

(Refer Slide Time: 00:22)



The word factorials is used when you go for experiment with several factors each having more than one levels. For example, assume there are 2 factors A and B, A with 2 levels and B with 2 levels. Pictorially we can represent like this, factor A with level 1 and level 2, factor B level 1 and level 2. So, then this is a factorial design with 2 factors and in this case you have 4 experimental settings independent experimental settings one is when factor A is at level 1, B at level 1 that is 1 1. Factor A at level 2, B at level 1 this is 2 1; factor A at level 1, B at level 2 that is 1 2 and then factor A and B are both are level 2 that is 2 2. Other way, otherwise also we write that this is low and this is high.

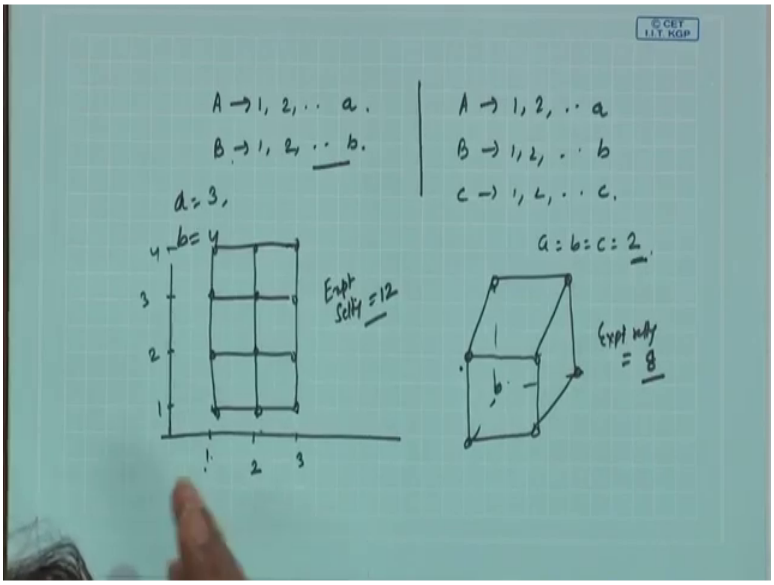
So, 1 stands for a values low value and B high value, similarly this one for similarly for factor B low and high value. For example, if factor A is temperature and we will find that 15 degree centigrade temperature is low and 25 degree centigrade temperature is high then we can put them as low high or 1 2 in instead of writing 15 and 25 degree centigrade. We can write like this, but more while you reach 1 2 low high sometimes you use minus 1 and plus 1 also. This is also low we write minus 1 and high we write plus 1 this is the notation is used. So, minus 1 and this will be plus 1.

Suppose the that factor B is reaction time then may be reaction time that is a low means may be 1 hour and high means may be 3 hours. So, what is low and what is high that is that will be determined by the expert because the operators the process expert if is a experiment related to any kind of process.

Now, here is a factorial design these design is factorial design having 2 levels each of the 2 factors A and B. So, there can be 3 factors there can be 4 factors and the level of the factors can be every factors may be more than 2 and that there is no had in first rule that a will be with A and B or other factors will be having equal number of levels. So, if this is the case then we have experimental settings or experimental and where we do the experimental experiment 1 2 3 4 and this is actually 1 2 1 2 3 and 4 giving like this, this is first setting, second setting, third setting, fourth setting. And now keeping A 1 B 1 means A at low, B at low, you can run the experiment 1 time, 2 times, 3 times n number of times then you will be getting response values depending on the replication 1 time, 2 time, 3 time or more number of times.

So, what is this y 111, y 112, y 113 all these are y values here, response values when you run the experiment here or similarly these, similarly these, similarly this, but it can be something like these A with 1 2 a levels, B also 1 2 b levels. So, there can be 3 factors A with 1 2 a levels, B with 1 2 b levels, C with 1 2 c levels.

(Refer Slide Time: 05:06)

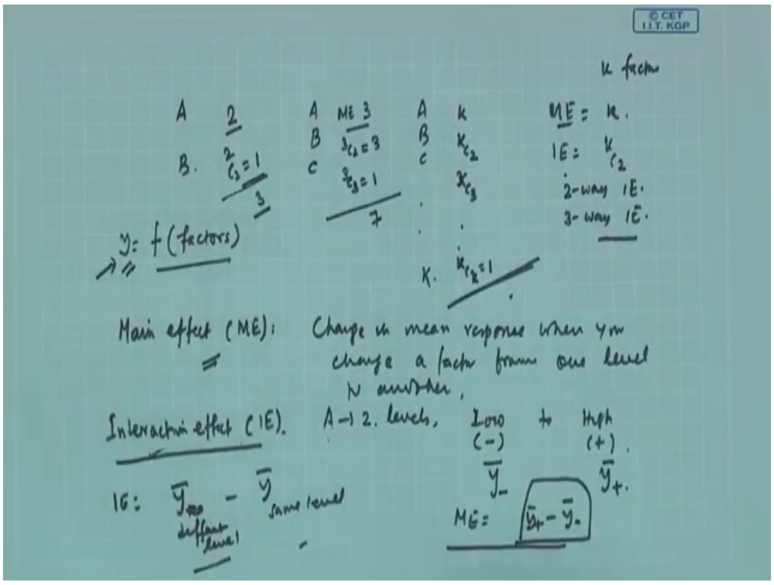


Suppose if a equal to b equal to c equal 2 then the experiment settings will be a cube. So, you have 8 points 8 experimental settings 1 2 3 4 5 6 7 8 and here and this is for simplification we have used this equal to 2. But suppose in this case suppose a equal to 3 and b equal to 4 then that the things will be like these settings will be a 1 2 3, b 1 2 3 4, then your number of experimental runs will be here this 1 2 3 then 1 2 3 4, 1 2 3 4, 1 2 3, so 1 2 3 4 5 6 7 8 9 10 11 12, so experimental settings here is 12.

Here experimental setting is equal to 8. So, these are known as treatment combination, especially treatment combinations. So, all those are basically example of factorial experiments and pictorial representation is possible when the number of levels as well as number of factors is 3 or 2 or less otherwise we have to use the tabular formulation.

So, what we will discuss today? We will discuss this under these factorial experiments, what are the key concepts and then what kind of data you will gather after experiment, and what are the hypothesis you will test here, and I will show you one example that hypothetical example for factorial experiment with data.

(Refer Slide Time: 08:17)



So, whatever may be the case if you have 2 factors, 3 factors, you can have factor 2 factors, you can have 3 factors, you can have k number of factors, k number of factors. So, few important things are that one is when we have several factors every factor effect the response y. So, y will be function of factors a b c d all those things. So, every factor might have influence on the observation y as well as the factors may be correlated or they might have interaction effects on y.

So, as a result the one of the important concept is called main effect. Main effect means this is basically the effect of changing the every one factor from one level to another level, primarily from low to high what will be the change in y that change mean main change. So, change in mean of y change in mean response when you change a factor from low to high from factor, from one level to another one level to another. For example if a is having 2 2 levels suppose what will be if you change from one level that low level which we say minus to high level which we may say plus. So, low to high then there will be change in y. So, this will be y bar minus, this may be y bar plus then what is the change? Change in average is y bar plus minus y bar minus this change is benefit.

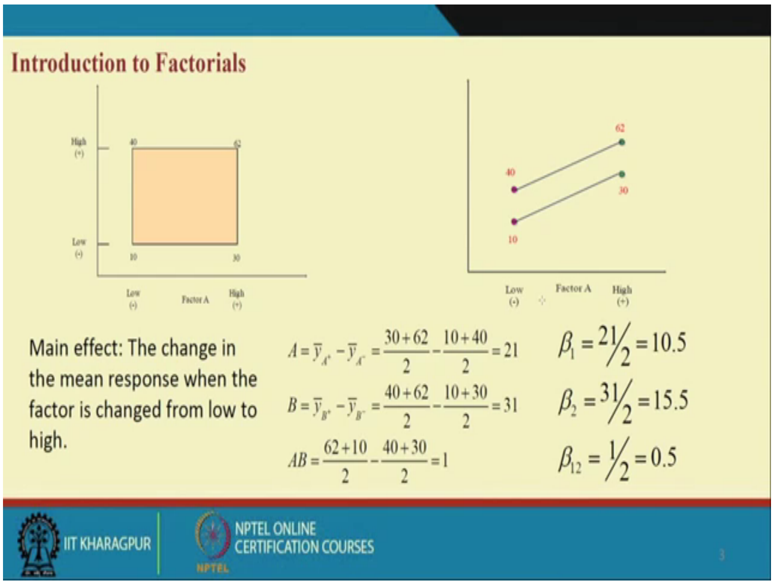
Another one will be the interaction effect. What is this interaction effect? It is something like this suppose you keep both the factors A and B at the same level 1 1 or 2 2 same level. If both low or both high whenever they are all same level and they will be at different level when A at low level B will be at high level, B at low A will be high level then what is the average change or change in mean response is when all are the factors are kept at that same level now with respect to the factors are that different levels. So, that mean the observation here and here will be averaged and observation here and here also will be averaged. So, the average of y these minus this will give you the interaction effect.

So, in that sense if I want to, that mean main effect something like this depending on whether the A or B, but it is basically what is the difference in mean when they are at high level at low level. So, it can be then interaction effect we can write like this y bar, may be both at high different level and suppose I say different level minus y bar same level. So, that will give you interaction effects.

So, another important thing is that then what will be the number of main effects, what will be the main effect total number of main effects. It will be, if there are k factors main effects will be k for example, here 2 factors, so 2 main effects. What do you mean by the interaction effects? Interaction effects will be k c 2 if there are 2 factors then it will be 2 C 2 will be 1. Similarly in this case 3 main effects and then you have another concept that in interaction effects one is 2 way interactions and there are 3 things every 2 factor will be interacting and all 3 factor will be interacting simultaneously, so 3 way interactions.

So, 2 way interaction mean 3 C 2 finally, it will be 3 and 3 way interaction 3 C 3 it will be 1. So, effectively 3 plus 3 plus 1, 7 effect you have to estimate it and here 3 effect you have to estimate. So, if you go for k number of factors, so main effect will be k 2 way will be k c 2 3 way will be k c 3. So, like these k c k last will be 1. So, it will be a large number of effects that is to be estimated. So, in addition, addition to this, please remember the replication will also be there in each experimental runs. So, the experimental settings it each every treatment combination there will be different experiment, replicated experiments. So, in that case you will be having more than one y value at each experimental settings or at each treatment combination you will be having n number of values. When we talk about averaging we basically ever take some of all including the replicated values and the other condition whether it is at low or high or at same level or different level and then we subtract.

(Refer Slide Time: 15:33)



Let us see one example. So, here we have 2 factors factor A and B, A at 2 levels, B at 2 levels it is denoted by low and high for or minus or plus. So, we are interested to compute the main effects and interaction effects. So, what you require? You require what are the values of response variable when you conduct experiment keeping at the factors at different levels.

So, when A at low and B at low then the y value what you are response value you observe is 10 and here we are considering one replication means single replicate n equal to 1. So, then when A at high B at low the response value is 30 and when A at low and B at high response value is 40 and when both are at high level response value is y value is 62. So, what do you want? We want to compute the main effects. So, suppose we are interested to compute the main effect of a factor A.

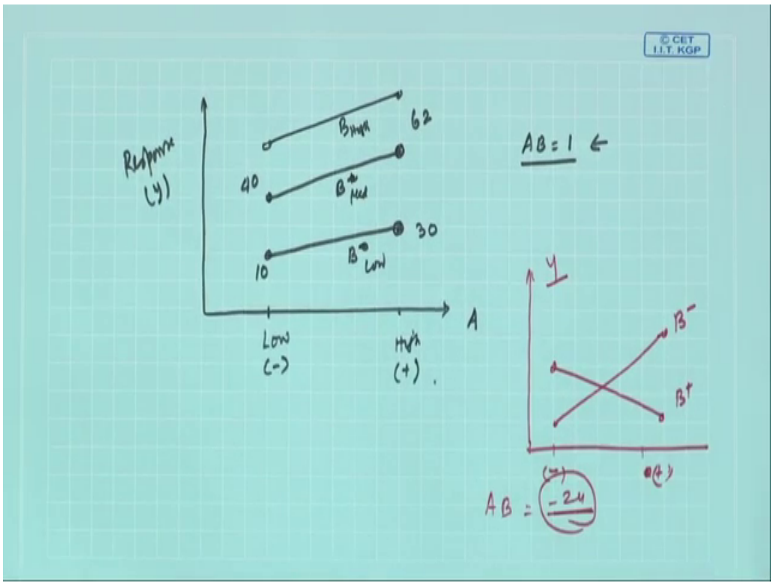
In that case what you do you take the average of response when A at high minus average of response when A at low. So, this is denoted by A equal to y, A plus minus A plus bar minus y A minus bar. So, then the value is thirteen plus 62 you see when a at high a at high value irrespective of the presence of B you consider when a at high what are the y values 30 and 62, 30 plus 62 by 2 because there are 2 observations 30 and 62. So, this is the average y a plus bar.

Similarly, when a at low that is 40 and 10 the average is 10 plus 40 by 2 the difference is giving you a value 21 this is what is the effect of a what does it mean if you change a from low to high y is a change will be 21. So, or if I say that it is from minus to plus means minus 1 to plus 1 the change is 21. In the same manner you can compute B effect. So, for B effect when B at high the average will be 62 plus 40 by 2 when B at low the average will be 30 plus 10 by 2. So, ultimately the difference is 31, it indicates that when you change B from minus 1 to plus 1 or low to high the change in y will be 31 units.

Now, come to the interaction, interaction A B here you see what is there A B. In a B case, you are basically considering when they are at high same level minus they are at different levels. So, 62 plus 10 by 2 that is why they are the same levels and then 40 plus 30 by 2 when at different levels, so 40 plus 30 by 2. So, that difference is 1. So, this A and B the value 21 and 31 and A B 1 these are the effect values A and B are the main effect A B is the interaction effect.

If you see the values you find out the A B values are quite high, but A B interaction is very low. So, it can be neglected, but later on we will see how to do the significant test of all those effects so that some of the effects will be neglected for further consideration. Now here, what happened? It is basically talking about, so you can see this graph I can write this graph.

(Refer Slide Time: 20:05)



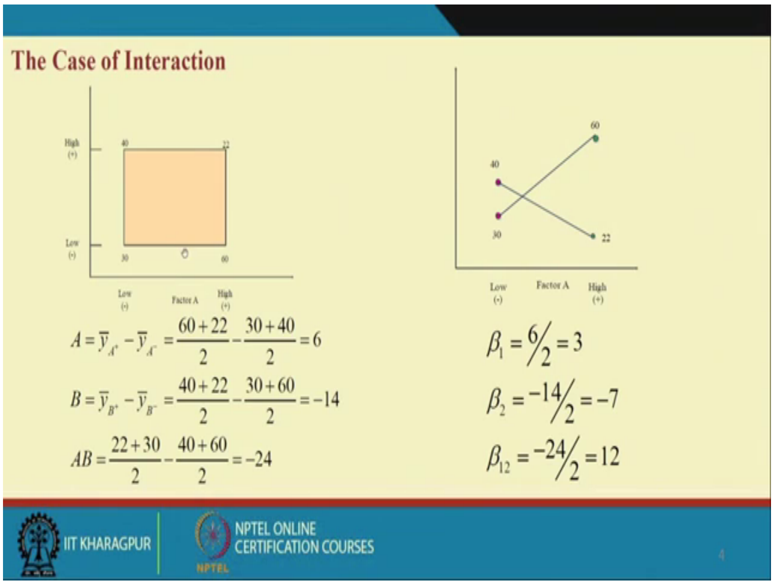
So, that the graph is response versus n suppose this side A and this side response. What is response? Response is y. So, A at low or minus to high or plus. Now consider B at low, when B at low A at low the value is 10, y value is 10, when A high and B low the value is

30, suppose this is 30. Then if I join I will get a line like this which is B minus; that means, this is A curve B at low, but you are changing A. Now, there will be similarly B plus curve show if this is 10, this is 30, let this is 40 when A at low B at high y response is 40 and when both are at high that is 62. So, something like this, this is 62 you do this. So, this is B positive.

Now, this two are more or less parallel and we have seen the A B interaction effect is one which is also very very less. So, if there is no interaction effect you will find out that when you plot the response values for a particular factor and then keeping another factor at a particular level you will get parallel lines. Now, if there is another level for, another level for B; that means, low medium and high then what will happen you also get may be similar another plot which is this is low, this is medium and this one is be high. So, there will be parallel lines. So, there is no interaction effects.

As we have seen that A if you make it from low to high the response is increasing and that is why this slope is coming.

(Refer Slide Time: 22:58)

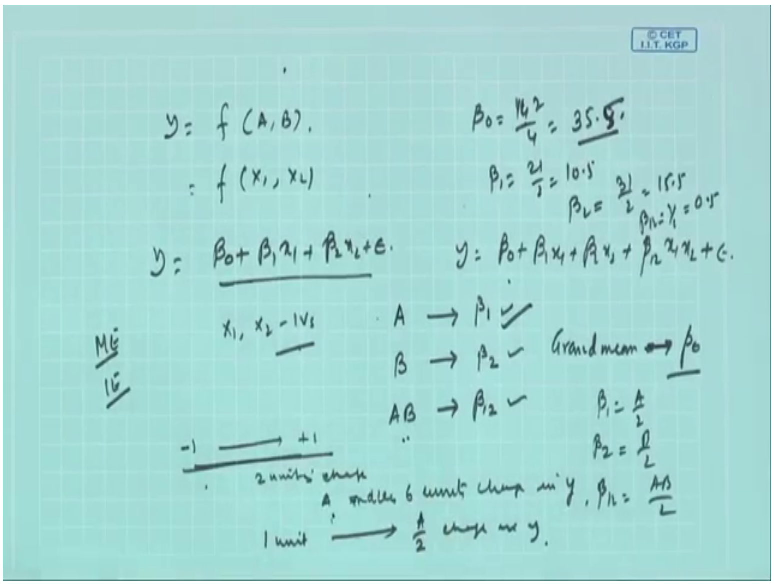


When there are interaction effects what will happen you see the next slide. In this case again the same that both two factors and you see that ultimately ultimately the response values are different here and when you compute the main effect a is 6 B is minus 14 and a B is minus 24.

So, what does it mean it means that A effect if you change from low to high then 6 unit change in response mean response for B it is minus 14 unit change in lowering and for A B interaction every in that combination will give you minus 24 units less. Now, if you this side A and this side response and then what happened it is with respect to B. So, when B at low this is 30 and 60 and when B at high 40 and 22. So that means, here when there is interaction what is happening when there is interaction you will get suppose A low to high you are making. So, you will find out something like this something like this.

So, the response values suppose this is B minus and this is B plus they will interact and then the A B value that will be y significant in this case it is minus 24, it is a high value. So, far then I have talked about 2 different factors and they are having a 2 levels minus and plus and so we can say y is function of A and B or in regression mode we can write it is x 1 and x 2. Then we can write down at this one is beta 0 plus beta 1 x 1 plus beta 2 x 2 plus epsilon.

(Refer Slide Time: 25:07)



Now, is there any relation between the main effects and interaction effects and the regression parameter. This is a regression equation when x 1 and x 2 are purely independent IVs.

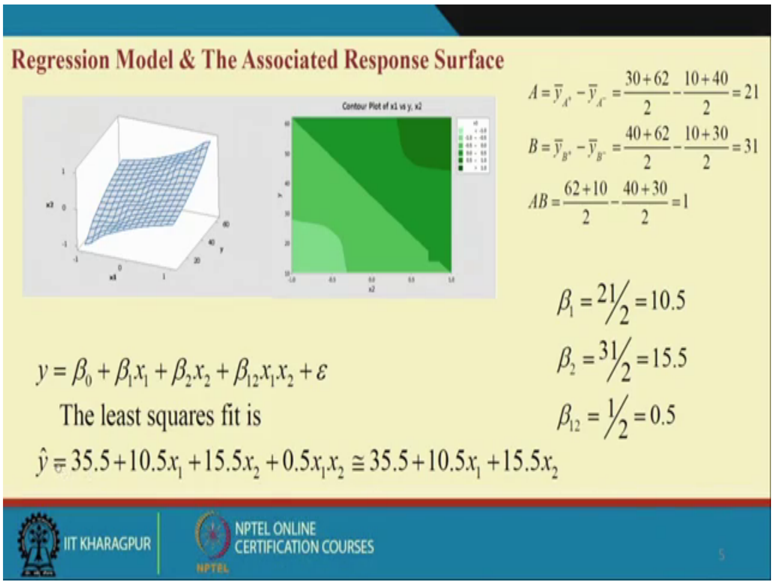
But if x 1 and x 2 are know, x 1 and x 2 have interaction effects what will happen this y will become beta 0 beta 1 x 1 beta 2 x 2 plus beta 1 2 x 1 x 2 plus epsilon. So, A is the main effect then it corresponding regression effect is beta 1, B is another main effect corresponding regression coefficient is beta 2 and then A B is interaction if it corresponding regression coefficient is beta 1 2. Then what is the beta 0? Beta 0 if the inter shape this is the grand mean. So, grand mean is actually represented by beta 0.

So, as you have seen that you have estimated A B, A B main effect and interaction if its can we get an estimate for the regression coefficient interestingly it is there. For example, if you take this case A is 6, but what is this? When you are changing from minus to plus, so minus 1 minus 1 we mean minus 1 to plus 1 you are making the change that mean 2 units change in A, 2 units change in A is represented by the effect change in A makes 6 units, units change in change in y. So, but when you compute regression coefficient that is one unit change in x 1 what is the change in y. So, that mean if one units change in y the x that will create; that means, A by 2 unit change in change in y. As a result beta 1 will be A by 2, beta 2 will be B by 2 and beta 1 2 will be A B by 2 that is what we have given here. That is what is given here.

And here beta 1 2 is very a large this is minus 12. Then here you see when other one that beta 1 is 10.5 and beta 2 is 15.5, beta 12 is 0.5 all are positives ok.

So, what will be the grand average here? 62 plus 40, 102 plus 40, so 142 divided by 4 that what will be the, that beta 0 value then for the this case, for the first example beta 0 will be 142 divided by 4. So, what is this? 4 into 3 22.5, 35.4. What is beta 1 value? Beta 1 is 21 by 2 that is 10.5 beta 2 will be 31 by 2 that is 15.5 and beta 1 2 will be 1 by 2 that is 0.5. Then what will be the result in regression equation?

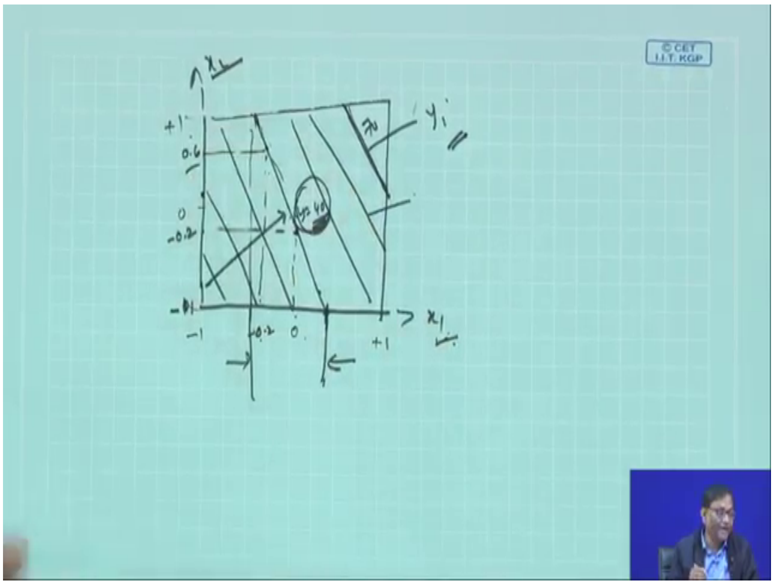
(Refer Slide Time: 29:13)



So, resultant this is 35.5 may be. So, yes 35.5 4 into 5. So, see the slide, what happen in the first example the regression equation will be y cap is the fitted y is 35.5 plus 10.5 x 1 15.5 x 2 plus 0.5 x 1 x 2. Now, this value is very less and later on if you find that that is not significant and then this can be this can be a reduced equation will be, like this so; that means, it is a linear regression plane and you will when you plot x 1 x 2 versus y you will find out a where play a basically a surface like this. So, here what happen x 1 changing from minus 1 to plus 1, x 2 changing from minus 1 to plus 1 then y is changing from different values from y that 22 to some value may be 10 2 when whatever the range is average. So, this is known as response surface.

Now, here in the right hand side contour plots given for that mean contour is that equal contour is the basically joining the points having equal property. For example, if you join on all the points on the half surface have been equal altitude that is from means level then this is a contour line. If you connect the cities with equal mean temperature annual mean temperature that is temperature contour line. In the same manner here what happen the contour line is we are basically joining the lines joining the points we with equal number or same value of y for different values of x. Essentially, you may get for this kind of when there is no interaction effect you will get this kind of lines this is suppose x 1 and this is x 2 and these are all y values. So, these are all y values, this is suppose y 1, this y 2 mean. What does it mean?

(Refer Slide Time: 31:38)

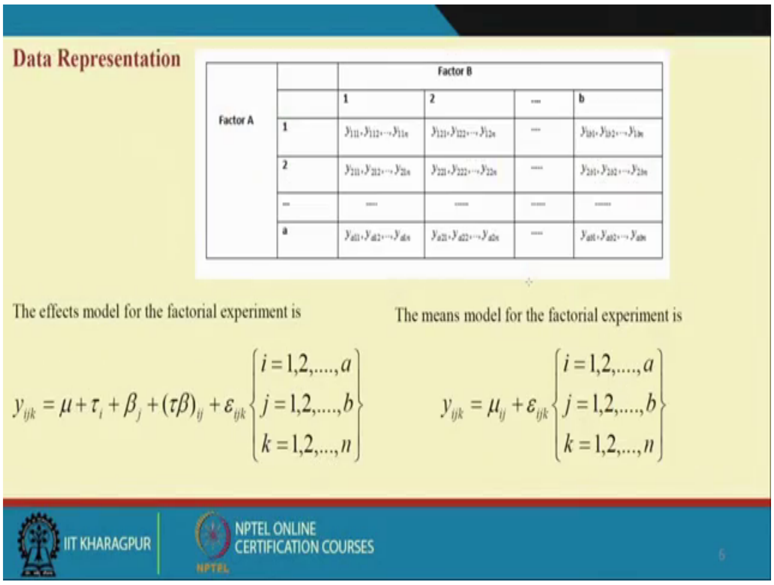


Suppose if I consider any point on this line is having a particular y value, suppose y is this will be 40 then all here also 40 here 40, here 40, everywhere 40 what does it mean if I consider this point and then make a projection here then x 1 this may be 0 this is minus 1 this is plus 1. So, this is again minus 1, this may be your 0, this is your plus 1 and if I do like these and this may be 0 and this is minus 1 this will be minus 0.2.

Suppose if I take a point here. So, this is again minus 0.2 and this may be 0.6. So, that mean if x 1 is the you have a range of our setting x 1 and x 2 or factor A and B, if you want y should be around 40 then this is the line contour line which gives you a range of x 1 and x 2 and so that you can set the machine or the process keeping different values of x 1 and different values of x 2, whether you that mean from here to this, here this is the range for x 1 and similarly this is the range for x 1 and similarly you will find out the range for that minus 1 to plus 1 this big range this is range for x 2.

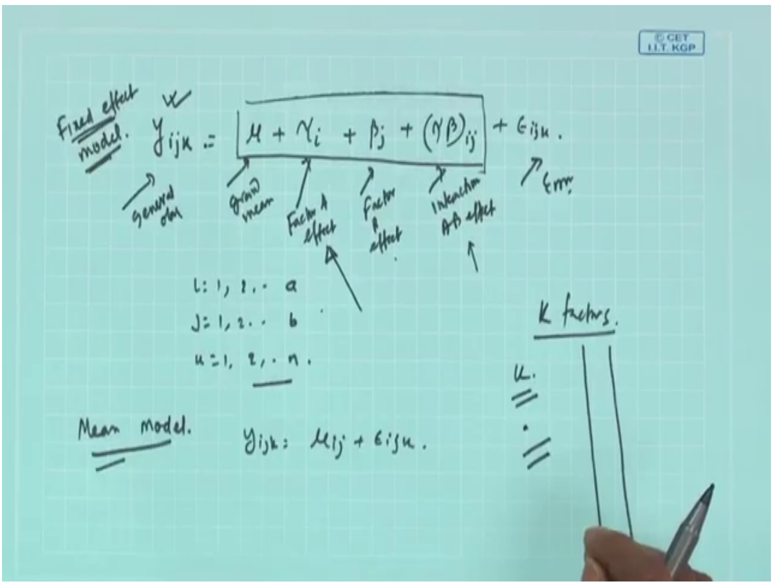
If you want y should be around 40 you choose any point here and corresponding x 1 x 2 will give you the proper parameter settings where you will fix. Suppose this is the line, suppose this one is 70 and your y is this line is 70 your y is maximum is a is a higher the better case then you will be interested to run the operation machine keeping x 1 may be one then y a x 2 will be something like this. So, this sends the surface response will be understood.

(Refer Slide Time: 34:28)



Now, I will give you from example to generalisation with a 2 factor case. So, the data representation for 2 factor factorial case will be factor A with A levels factor B with B levels, then and if there are n replication then your data set will be like this. And what will be the statistical model? Statistical model will be something like.

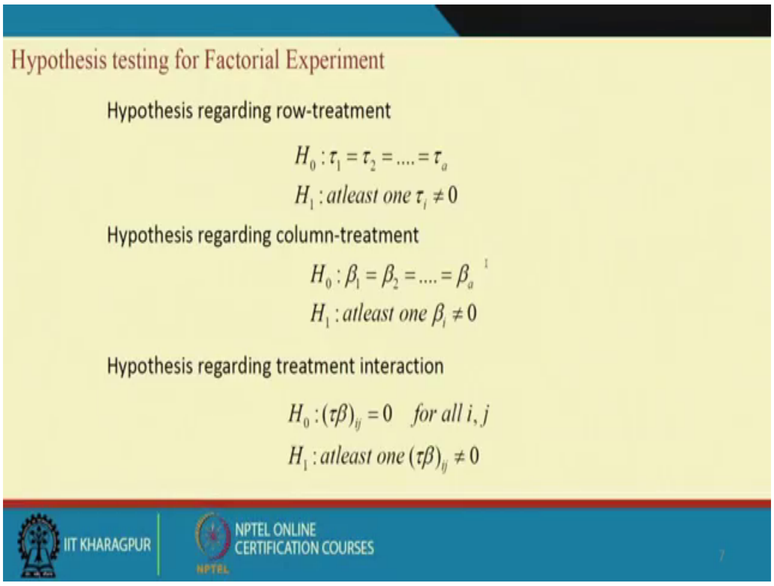
(Refer Slide Time: 35:02)



This we have y i for factor A j for factor B and k for replication and then this will be having a grand mean plus tau I that is factor A effect beta j, factor B effect plus tau beta ij that is the interaction effect plus epsilon ijk. So, this is the general observation, this is grand mean, this is our factor A effect, factor B effect, this is interaction A B effect, this is your error, and i stands from 1 to a, j stands from 1 to b, k 1 to n. This is known as fixed effect model.

Fixed effect model you may be interested in a mean model if you are interested in a mean model then what happen the entire thing this will be written as mu ij and then y ijk will be mu ij plus epsilon ijk, this is known as mean model. So, if I say for 2 factorial experiment what is the statistical model, fixed effect model is this what will be your role your role will be to estimate all the parameters and then there after you test whether this parameter the this model is significantly contributing in explaining the y variability or not. So, that we will see later on and accordingly we test certain amount of hypothesis, anyhow.

(Refer Slide Time: 37:22)

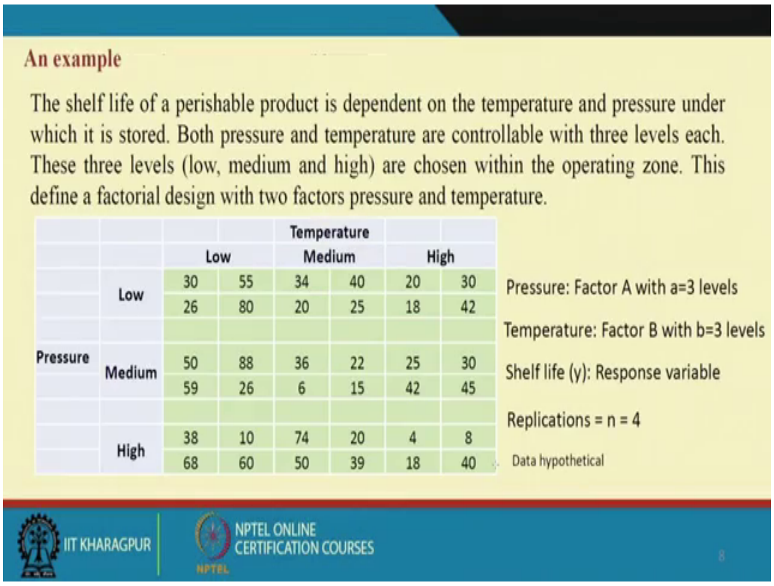


So, for the time being we just will later see the x hypothesis also. So, what happen? What would will be the hypothesis? Hypothesis will be whether the main effects are significant individual main effect significant or not, individual interaction effect significant or not. In this case we have 2 treatment, one is row treatment, another one is column treatment that is factor A and factor B and also we have interaction AB.

So, what we are writing H 0 that there is no difference between the treatment levels row treatment levels and H 1 at least there is one difference. Similarly for column treatment level there is no difference H 1 there is difference and hypothesis regarding treatment interactions. So, there is the treatment effects are 0 or interaction effects are 0 for all ij at least one tau beta ij not equal to 0. So, these are the hypothesis.

So, we are basically we want to test here 3 kinds of hypothesis, one for factor A, another for factor B and another for interaction AB. So, if you have k number of factors, k factors. So, how many test you will be doing for all main effects k main effects, 2 way interaction effects, 3 way interaction effects. So, those many effects are there all main an interaction effects including 2 way to multi way k 2 way interaction effect all those effects for all those effects you will be basically putting one hypothesis. If there are k factors k hypothesis for main effects and then k c 2 hypothesis for 2 way interaction effects k c 3 hypothesis for 3 way interaction effects like this you have to do.

(Refer Slide Time: 39:42)



So, we will see one experiment in the next class and all those things what we have discussed so far, how the factorial experiments is developed and all those things. Suppose if I consider this experiment that the shelf life of a perishable product is dependent on temperature and pressure under which it is stored. So, there are 2 controllable factors temperature and pressure which basically governs the shelf life. Both and both pressure and temperature are controllable with 3 levels each. So, low medium high and low medium high. These 3 levels are chosen within the operating zone it is basically it must be in the operating zone. This define a factorial design with 2 factors pressure this is a factorial design.

Now, let us assume that you have done the experiment and basically you have seen what is actually happening in the store and then suppose when it is temperature pressure are low level low then you have 4 replication for low medium 4 for low high 4 similarly medium low 4. So, 4 replications then this is a factorial experiment with 2 factors factor A at 3 levels, factor B at 3 levels shelf life is the response variable and replication is 4 here the data is hypothetical.

So, when you do factorial experiment depending on the problem and depending on the number of controllable factors that is to be controlled and depending on the sample size requirement you will be having the upper experiment you will be having data similar to this.

So, I hope you understand the factorial experiment and also you know what will be the statistical model for factorial experiment. In next class we will be discussing on how to see that whether the model is a able to explain the variability of the data using ANOVA and then next to that we will see that how the different parameters of the model will be estimated, and followed by the residual analysis or the test of assumptions and other things.

Thank you very much.