

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)



Basic Definitions and Principles

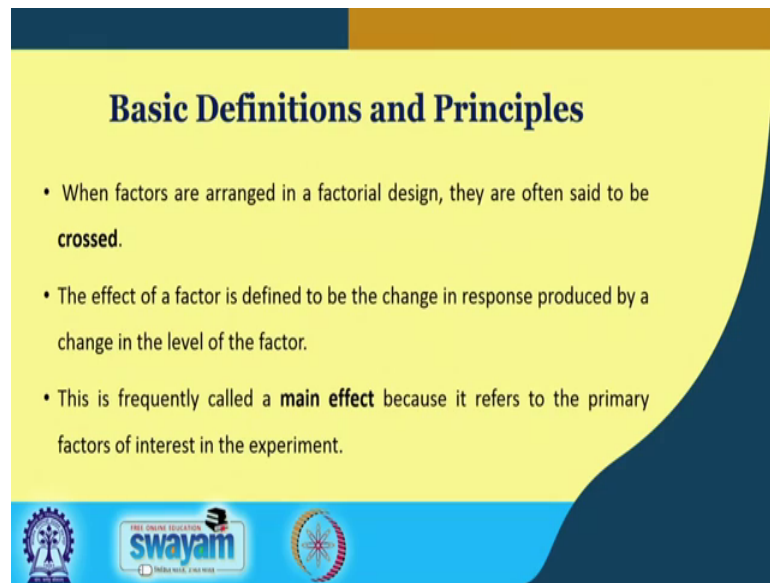
- Many experiments involve the study of the effects of two or more factors.
- In general, **factorial designs** are most efficient for this type of experiment.
- By a factorial design, we mean that in each complete trial or replicate of the experiment all possible combinations of the levels of the factors are investigated.
- For example, if there are a levels of factor A and b levels of factor B, each replicate contains all ab treatment combinations.

Logos at the bottom: IIT Bombay, SWAYAM, and a circular emblem.

So, the basic definition in principle is like this that factorial designs are most efficient for this type of experimentation when you have to analyse two or more factors. Factorial design means that in each complete trial or replicate of the experiment all possible combinations 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)



Basic Definitions and Principles

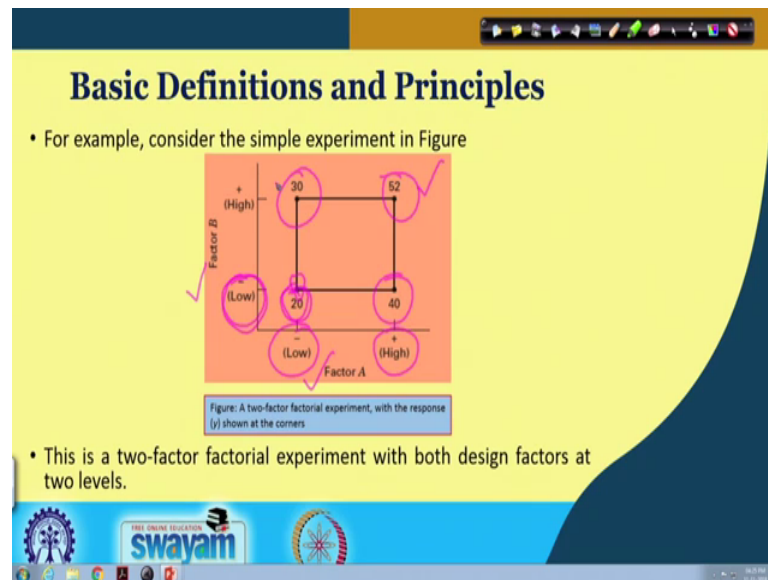
- When factors are arranged in a factorial design, they are often said to be **crossed**.
- The effect of a factor is defined to be the change in response produced by a change in the level of the factor.
- This is frequently called a **main effect** because it refers to the primary factors of interest in the experiment.

Logos at the bottom: IIT Bombay, SWAYAM, and a circular emblem.

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)

Basic Definitions and Principles

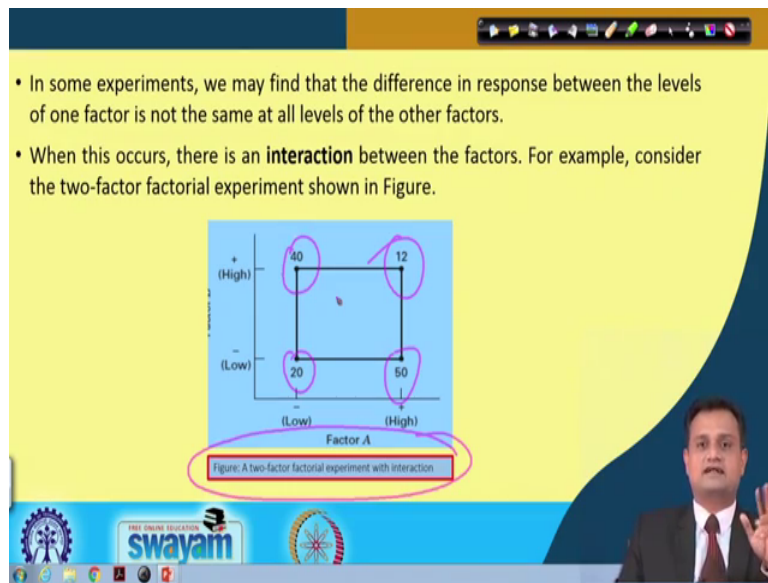
- The main effect of factor A in this two-level design can be thought of as the difference between the average response at the low level of A and the average response at the high level of A.
- Numerically, this is
$$A = \frac{40 + 52}{2} - \frac{20 + 30}{2} = 21$$
- That is, increasing factor A from the low level to the high level causes an **average response increase** of 21 units.
- Similarly, the main effect of B is
$$B = \frac{30 + 52}{2} - \frac{20 + 40}{2} = 11$$

swayam

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.

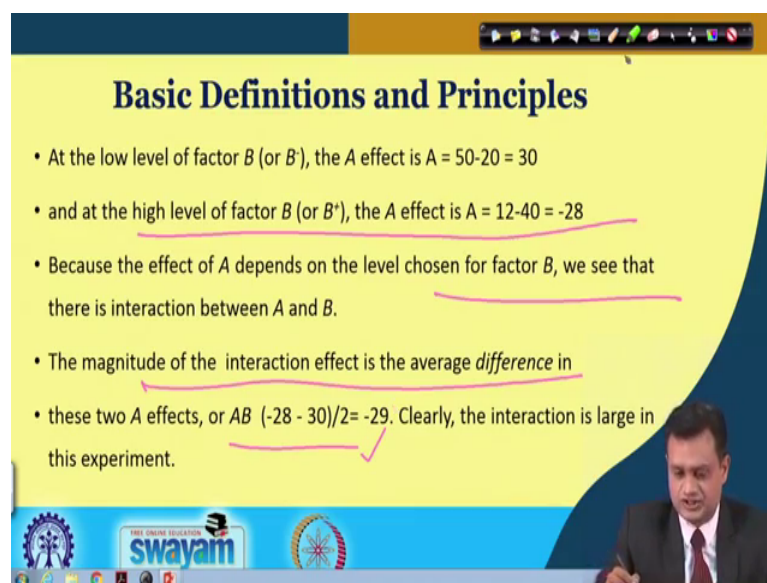
(Refer Slide Time: 07:00)



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)



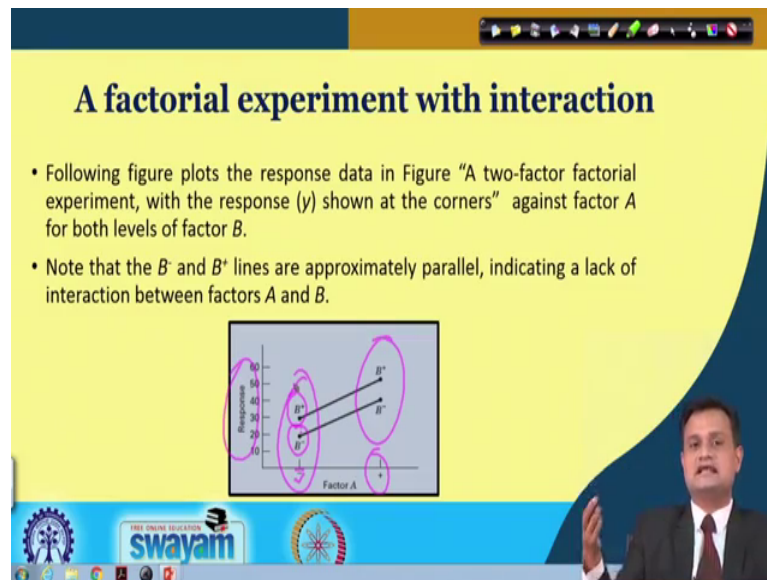
Basic Definitions and Principles

- At the low level of factor B (or B^-), the A effect is $A = 50 - 20 = 30$
- and at the high level of factor B (or B^+), the A effect is $A = 12 - 40 = -28$
- Because the effect of A depends on the level chosen for factor B , we see that there is interaction between A and B .
- The magnitude of the interaction effect is the average difference in
- these two A effects, or $AB = (-28 - 30)/2 = -29$. Clearly, the interaction is large in this experiment.

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)



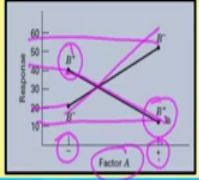
Now, just see that when you have factorial experiment with interaction, then a two factor of factorial experiment with a response y you can show it as let us say corners and this is my response against factor A. So, against factor A when it is at low level, what is happening in this? So, B plus and B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

A factorial experiment without interaction

- Similarly, Following figure plots the response data in Figure "A two-factor factorial experiment with interaction."
- Here we see that the B^- and B^+ lines are not parallel.
- This indicates an interaction between factors A and B.
- Two-factor interaction graphs such as these are frequently very useful in interpreting significant interactions and in reporting results to non statistically trained personnel.



The graph shows a two-factor factorial experiment with interaction. The y-axis is labeled 'Response' and ranges from 10 to 60. The x-axis is labeled 'Factor A' and has two levels: '-' and '+'. There are two lines: a solid line for B+ and a dashed line for B-. The B+ line starts at approximately 25 for Factor A '-' and rises to approximately 55 for Factor A '+'. The B- line starts at approximately 45 for Factor A '-' and falls to approximately 20 for Factor A '+'. The lines are not parallel, indicating an interaction. The graph is annotated with pink circles and arrows highlighting the non-parallel nature of the lines.


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)

Basic Definitions and Principles

- Suppose that both of our design factors are **quantitative** (such as temperature, pressure, time, etc.).
- Then a **regression model representation** of the two-factor factorial experiment could be written as
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$
- where y is the response, the β 's are parameters whose values are to be determined, x_1 is a variable that represents factor A , x_2 is a variable that represents factor B , and ϵ is a random error term.
- The variables x_1 and x_2 are defined on a **coded scale** from -1 to 1 (the low and high levels of A and B), and $x_1 x_2$ represents the interaction between x_1 and x_2 .

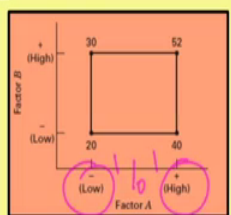


And then my statistical model would look like this where y is equal to β_0 that is the intercept. You have $\beta_1 x_1$ that captures the main effect, $\beta_2 x_2$ that captures another factor effect, main effect, $\beta_{12} x_1 x_2$ and you have ϵ that is your error component. So, this is how my model will look like.

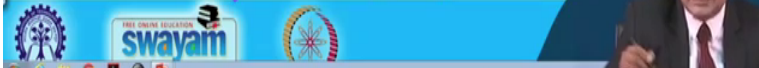
(Refer Slide Time: 12:34)

Basic Definitions and Principles

- For the experiment shown in Figure, we found the main effects of A and B to be $A=21$ and $B=11$.



- The estimates of β_1 and β_2 are one-half the value of the corresponding main effect; therefore, $\hat{\beta}_1 = \frac{21}{2} = 10.5$ and $\hat{\beta}_2 = \frac{11}{2} = 5.5$.



Now, let us try to go little bit deeper and investigate the values of the slope. So, what I will say that β_1 , this is the system I have already explained and β_1 is basically estimate $\hat{\beta}_1$ is 21 by 2 and β_2 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 β_1 , this will be β_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)

Basic Definitions and Principles

- The interaction effect in Figure is $AB=1$, so the value of interaction coefficient in the regression model is $\hat{\beta}_{12} = \frac{1}{2} = 0.5$.
- The parameter β_0 is estimated by the average of all four responses, or $\beta_0 = (20+40+30+52)/4=35.5$.
- Therefore, the fitted regression model is

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2$$

But I can estimate my β_1 β_2 and same way I can do it $\beta_1 \beta_2$. So, in figure A, B is equal to 1, so, I just divided by 1 by 2 and I have β_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 \times 1 plus 5.5, \times 2 plus 0.5, \times 1 into \times 2.

(Refer Slide Time: 15:06)

- Following figure presents graphical representations of this model.
- In Figure (a) we have a plot of the plane of y -values generated by the various combinations of x_1 and x_2 . This three dimensional graph is called a **response surface plot**.
- Figure (b) shows the contour lines of constant response y in the x_1, x_2 plane.
- Notice that because the response surface is a plane, the contour plot contains parallel straight lines.

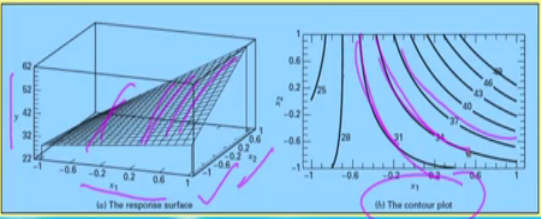


Figure (a) The response surface and Figure (b) The contour plot.

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)

- Now suppose that the interaction contribution to this experiment was not negligible;
- that is, the coefficient β_{12} was not small.
- Following figure presents the response surface and contour plot for the model.
- $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$
- Notice that the significant interaction effect "twists" the plane in Figure a (in the next slide).
- This twisting of the response surface results in curved contour lines of constant response in the x_1, x_2 plane, as shown in Figure b (in the next slide).
- Thus, **interaction is a form of curvature** in the underlying response surface model for the experiment.

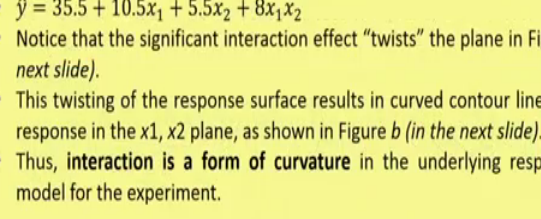
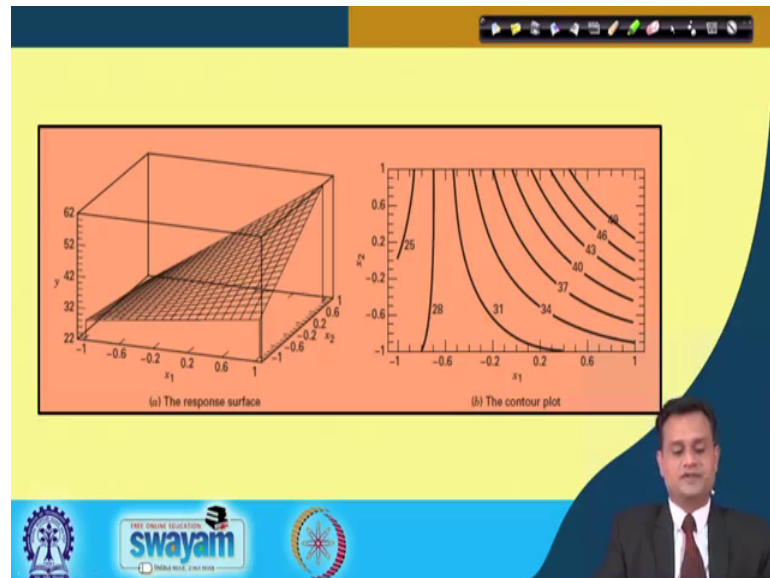


Figure (a) The response surface and Figure (b) The contour plot.

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)

- Suppose we have two factors A and B, each at two levels. We denote the levels of the factors by A^+ , A^- , B^+ , and B^- .
- Information on both factors could be obtained by varying the factors one at a time, as shown in Figure.
- The effect of changing factor A is given by $A^+B^- - A^-B^-$, and the effect of changing factor B is given by $A^+B^+ - A^-B^+$.
- Because experimental error is present, it is desirable to take two observations, say, at each treatment combination and estimate the effects of the factors using average responses. Thus, a total of six observations are required.

1.C. A B
+
+
+
+
+
+

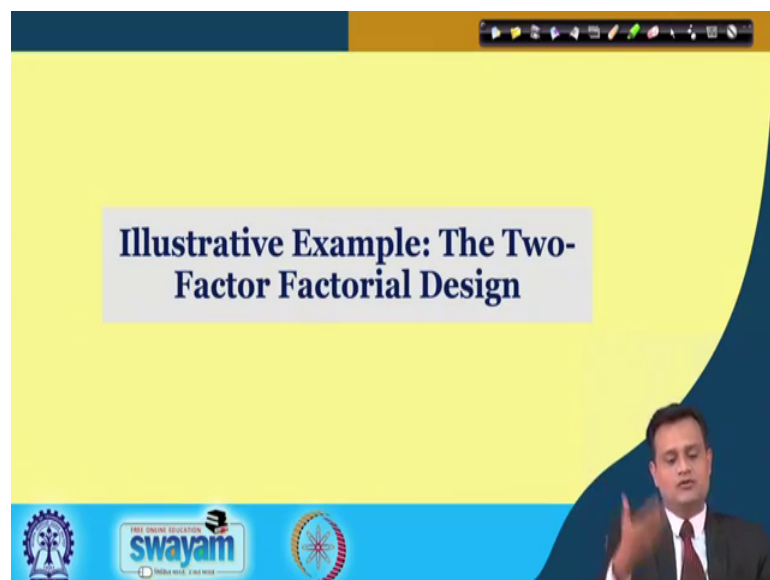
One Factor-at-a-Time (OFAT) experiment

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)

• Because there are two factors at three levels, this design is sometimes called a **32 factorial design**.

• Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order.

• In this problem the engineer wants to answer the following questions:

1. What effects do material type and temperature have on the life of the battery?
2. Is there a choice of material that would give *uniformly long life regardless of temperature*?

Material Type	Temperature (F)					
	15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

swayam

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)

• In general, a two-factor factorial experiment will appear as in follow Table: General Arrangement for a Two-Factor Factorial Design

		Factor B			
		1	2	...	b
Factor A	1	$y_{111}, y_{112}, \dots, y_{11n}$	$y_{121}, y_{122}, \dots, y_{12n}$		$y_{1b1}, y_{1b2}, \dots, y_{1bn}$
	2	$y_{211}, y_{212}, \dots, y_{21n}$	$y_{221}, y_{222}, \dots, y_{22n}$		$y_{2b1}, y_{2b2}, \dots, y_{2bn}$
	.				
	a	$y_{a11}, y_{a12}, \dots, y_{a1n}$	$y_{a21}, y_{a22}, \dots, y_{a2n}$		$y_{ab1}, y_{ab2}, \dots, y_{abn}$

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)

• The order in which the abn observations are taken is selected at random so that this design is a **completely randomized design**.

• The observations in a factorial experiment can be described by a model. There are several ways to write the model for a factorial experiment. The **effects model** is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

Where

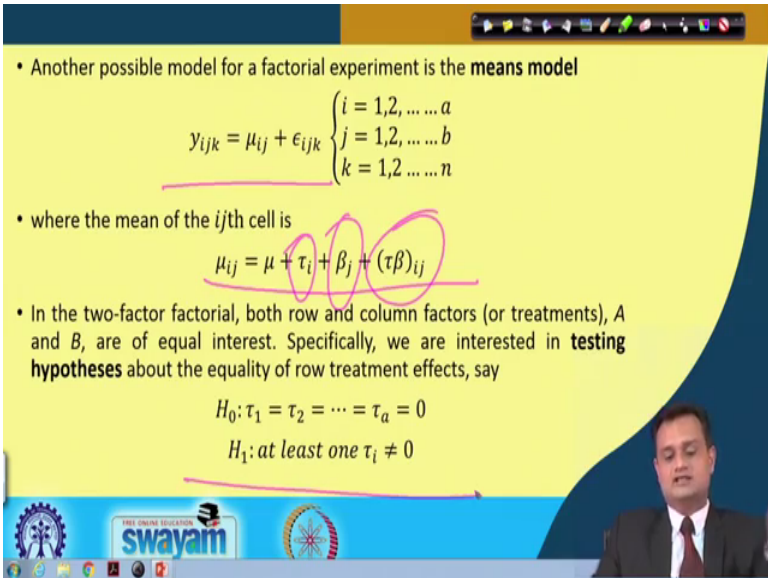
- μ = overall mean
- τ_i = effect of the i th level of the row factor A
- β_j = effect of the j th level of the column factor B
- $(\tau\beta)_{ij}$ = effect of the interaction between τ_i and β_j
- ϵ_{ijk} = random error component

But before that let me try to understand my model, what is my factorial design model? So, my model looks like this y_{ijk} this is the effect model I am trying to analyse the

effect, y_{ijk} 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, μ is the overall mean, τ_i is the effect of the level of the row factor A , then β_j is the effect of the j th level of the column factor B $\tau_i\beta_j$ is the effect of the interaction between τ_i and β_j and this is your random error component that is ϵ_{ijk} .

(Refer Slide Time: 21:19)



- Another possible model for a factorial experiment is the **means model**

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

- where the mean of the ij th cell is

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$$

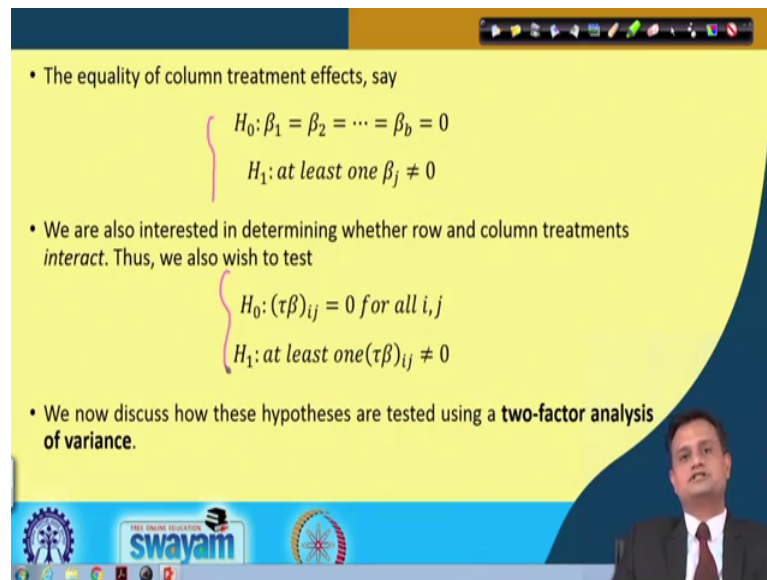
- In the two-factor factorial, both row and column factors (or treatments), A and B , are of equal interest. Specifically, we are interested in **testing hypotheses** about the equality of row treatment effects, say

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$$

$$H_1: \text{at least one } \tau_i \neq 0$$

So, when you appreciate this basic model what I say that y_{ijk} basically is the mean specific to i and j plus ϵ_{ijk} and μ_{ij} . The way its represented in a statistical mathematical term μ plus there is τ_i effect, row effect, column effect, interaction effect and I can have a null hypothesis that τ_i is equal to 0 means τ_1 is equal to τ_2 is equal to τ_a is equal to 0 and I can say that at least one of the treatment τ_a or τ_i is not equal to 0.

(Refer Slide Time: 22:04)



• The equality of column treatment effects, say

$$\left\{ \begin{array}{l} H_0: \beta_1 = \beta_2 = \dots = \beta_b = 0 \\ H_1: \text{at least one } \beta_j \neq 0 \end{array} \right.$$

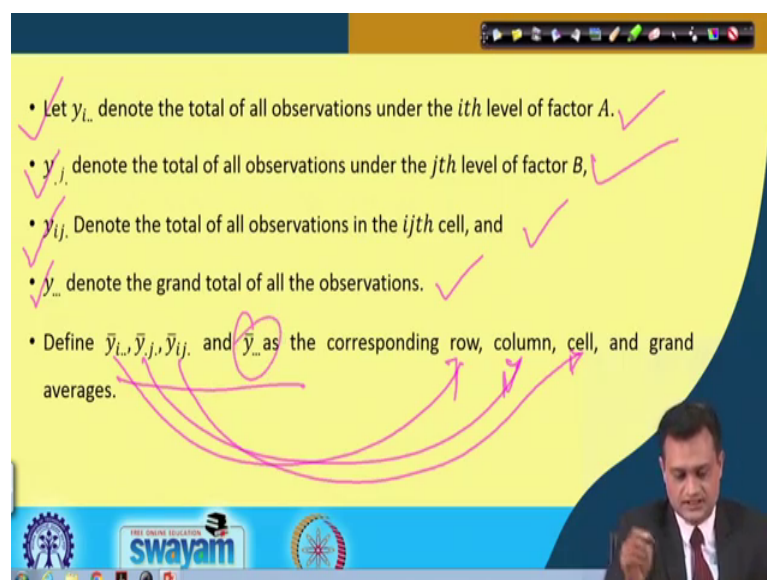
• We are also interested in determining whether row and column treatments *interact*. Thus, we also wish to test

$$\left\{ \begin{array}{l} H_0: (\tau\beta)_{ij} = 0 \text{ for all } i, j \\ H_1: \text{at least one } (\tau\beta)_{ij} \neq 0 \end{array} \right.$$

• We now discuss how these hypotheses are tested using a **two-factor analysis of variance**.

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)



• Let $y_{i..}$ denote the total of all observations under the i th level of factor A. ✓

• $y_{.j.}$ denote the total of all observations under the j th level of factor B, ✓

• $y_{ij.}$ Denote the total of all observations in the ij th cell, and ✓

• $y_{...}$ denote the grand total of all the observations. ✓

• Define $\bar{y}_{i..}$, $\bar{y}_{.j.}$, $\bar{y}_{ij.}$ and $\bar{y}_{...}$ as the corresponding row, column, cell, and grand averages. ✓

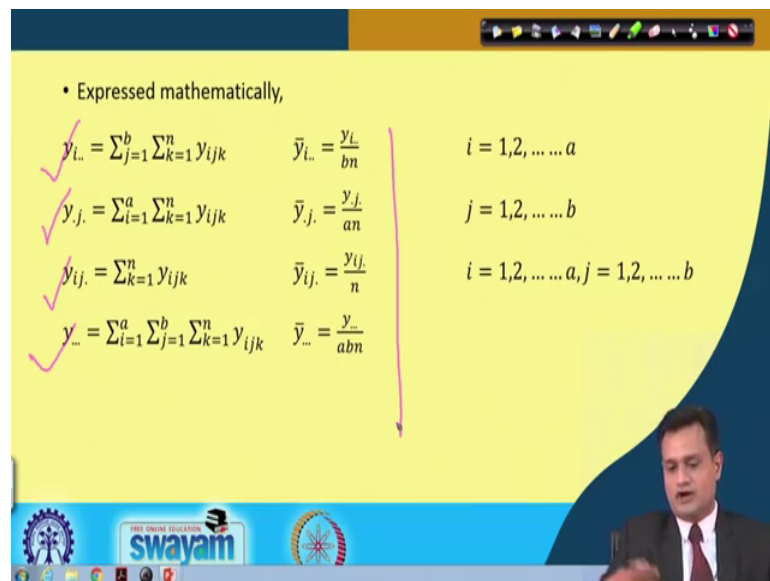
Handwritten pink arrows connect the averages to their respective totals: $\bar{y}_{i..}$ to $y_{i..}$, $\bar{y}_{.j.}$ to $y_{.j.}$, $\bar{y}_{ij.}$ to $y_{ij.}$, and $\bar{y}_{...}$ to $y_{...}$.

So, statistical analysis of the model is like this you have $y_{i..}$ triple dot basically denote the total of all the observations. You take the summation under the i th level of factor A, you

have $y_{.j}$ denote the total of all the observations under j th level of factor B, you have $y_{i.}$ denote the total of all the observation in say particularly i j th cell and $y_{...}$ denote the grand total of all the observations.

So, these are the mean values corresponding mean values $y_{i.}$ refers to row $y_{.j}$ dot refers to column y_{ij} dot bar refers to cell and then, you have the grand averages.

(Refer Slide Time: 23:20)



• Expressed mathematically,

$y_{i.} = \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$	$\bar{y}_{i.} = \frac{y_{i.}}{bn}$	$i = 1, 2, \dots, a$
$y_{.j} = \sum_{i=1}^a \sum_{k=1}^n y_{ijk}$	$\bar{y}_{.j} = \frac{y_{.j}}{an}$	$j = 1, 2, \dots, b$
$y_{ij.} = \sum_{k=1}^n y_{ijk}$	$\bar{y}_{ij.} = \frac{y_{ij.}}{n}$	$i = 1, 2, \dots, a, j = 1, 2, \dots, b$
$y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$	$\bar{y}_{...} = \frac{y_{...}}{abn}$	

So, these are the values you need to compute and then you can say find the total of $y_{i.}$ double dot $y_{.j}$ dot using this expression $y_{ij.}$ dot and $y_{...}$ triple dot and this is the averages of each particular $y_{i.}$ double dot bar $y_{.j}$ 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)

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \left[(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.}) \right]^2 \\ &= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &\quad + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \\ &\quad + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 \end{aligned}$$

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_{ijk} minus $\bar{y}_{...}$. 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)

- From the last component on the right-hand side of Equation, we see that there must be at least two replicates ($n \geq 2$) to obtain an error sum of squares.
- We may write Equation symbolically as

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$
- The number of degrees of freedom associated with each sum of squares is

Effect	Degree of Freedom
A	$(a - 1)$
B	$(b - 1)$
AB interaction	$(a - 1)(b - 1)$
Error	$ab(n - 1)$
Total	$abn - 1$

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)

- Each sum of squares divided by its degrees of freedom is a **mean square**. The expected values of the mean squares are

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \sigma^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a-1}$$

$$E(MS_B) = E\left(\frac{SS_B}{b-1}\right) = \sigma^2 + \frac{an \sum_{j=1}^b \beta_j^2}{b-1}$$

$$E(MS_{AB}) = E\left(\frac{SS_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + \frac{n \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

and

$$E(MS_E) = E\left(\frac{SS_E}{ab(n-1)}\right) = \sigma^2$$

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)

Source of variation	Sum of squares	Degree of freedom	Mean squares	F_0 Value
A treatment	SS_A	$a - 1$	$\frac{SS_A}{a - 1}$	$F_0 = \frac{MS_A}{MS_E}$
B treatment	SS_B	$b - 1$	$\frac{SS_B}{b - 1}$	$F_0 = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a - 1)(b - 1)$	$\frac{SS_{AB}}{(a - 1)(b - 1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n - 1)$	$\frac{SS_E}{ab(n - 1)}$	
Total	SS_T	$abn - 1$		

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)

- Computing formulas in terms of row, column, and cell totals can also be used. The total sum of squares is computed as usual by

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

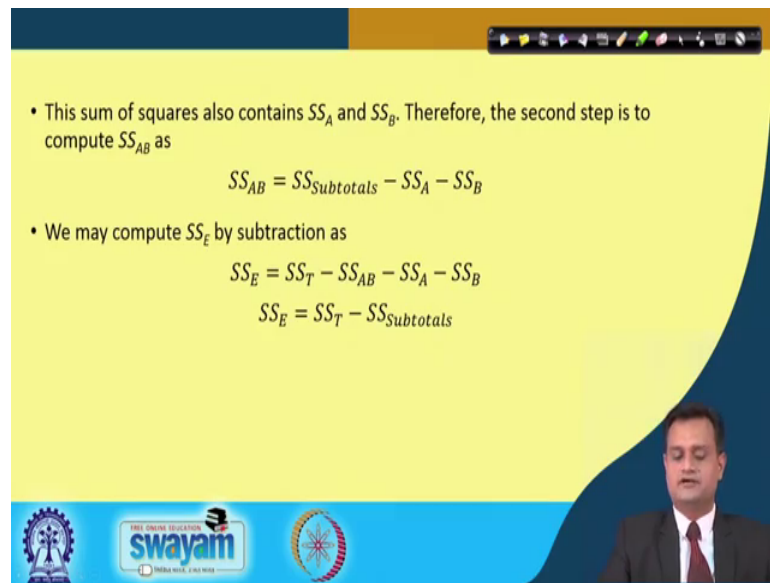
- The sums of squares for the main effects are

$$SS_A = \frac{1}{bn} \sum_{i=1}^a y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$SS_B = \frac{1}{an} \sum_{j=1}^b y_{.j.}^2 - \frac{y_{...}^2}{abn}$$

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)



This slide contains two bullet points and two equations. The first bullet point states: "This sum of squares also contains SS_A and SS_B . Therefore, the second step is to compute SS_{AB} as". Below this is the equation $SS_{AB} = SS_{Subtotals} - SS_A - SS_B$. The second bullet point states: "We may compute SS_E by subtraction as". Below this are two equations: $SS_E = SS_T - SS_{AB} - SS_A - SS_B$ and $SS_E = SS_T - SS_{Subtotals}$. The slide features a yellow background with a blue and orange header. At the bottom, there are logos for "swayam" and "INDIA WIDE, 24x7 WIDE" along with a small video feed of a man in a suit.

- This sum of squares also contains SS_A and SS_B . Therefore, the second step is to compute SS_{AB} as
$$SS_{AB} = SS_{Subtotals} - SS_A - SS_B$$
- We may compute SS_E by subtraction as
$$SS_E = SS_T - SS_{AB} - SS_A - SS_B$$
$$SS_E = SS_T - SS_{Subtotals}$$

(Refer Slide Time: 27:05)



This slide features a central grey box with the text "Illustrative Example: Battery Design Experiment". The slide has a yellow background with a blue and orange header. At the bottom, there are logos for "swayam" and "INDIA WIDE, 24x7 WIDE" along with a small video feed of a man in a suit.

**Illustrative Example:
Battery Design Experiment**

(Refer Slide Time: 27:08)

Table presents the effective life (in hours) observed in the battery design example. The row and column totals are shown in the margins of the table, and the circled numbers are the cell totals.

Life Data (in hours) for the Battery Design Experiment										
Material Type	Temperature (F)									$y_{i.}$
	15			70			125			
1	130	155	539	34	40	229	20	70	230	998
	74	180		80	75		82	58		
2	150	188	623	136	122	479	25	70	198	1300
	159	126		106	115		58	45		
3	138	110	376	174	120	583	96	104	342	1501
	168	160		150	139		82	60		
$y_{.j}$	1738			1291			770			3799 = $y_{..}$

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)

• Using the Equations mentioned previously, the sums of squares are computed as follows:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{..}^2}{abn}$$

$$= (130)^2 + (155)^2 + (70)^2 + \dots + (60)^2 - \frac{(3799)^2}{36} = 77,646.97$$

$$SS_{Material} = \frac{1}{bn} \sum_{i=1}^a y_{i.}^2 - \frac{y_{..}^2}{abn}$$

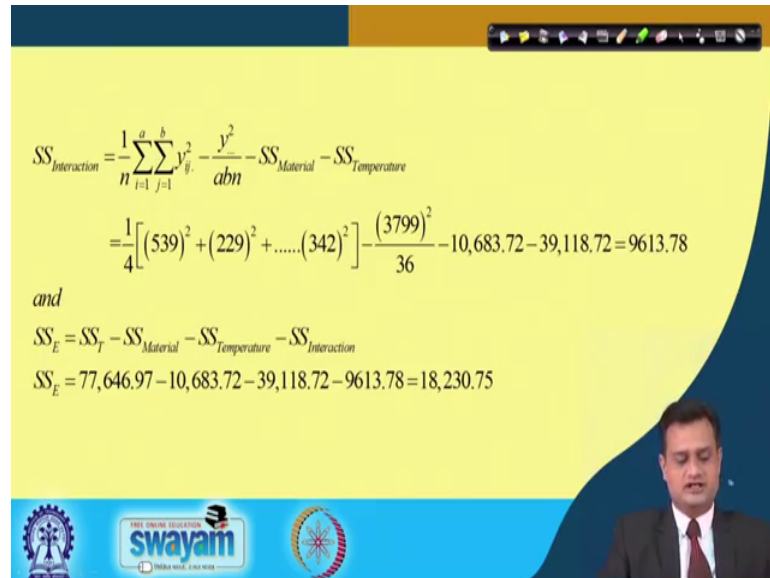
$$= \frac{1}{(3)(4)} [(998)^2 + (1300)^2 + (1501)^2] - \frac{(3799)^2}{36} = 10,683.72$$

$$SS_B = \frac{1}{an} \sum_{j=1}^b y_{.j}^2 - \frac{y_{..}^2}{abn}$$

$$= \frac{1}{(3)(4)} [(1738)^2 + (1291)^2 + (770)^2] - \frac{(3799)^2}{36} = 39118.72$$

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)



Slide content showing the calculation of SS Interaction and SS Error:

$$SS_{Interaction} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y_{..}^2}{abn} - SS_{Material} - SS_{Temperature}$$

$$= \frac{1}{4} \left[(539)^2 + (229)^2 + \dots + (342)^2 \right] - \frac{(3799)^2}{36} - 10,683.72 - 39,118.72 = 9613.78$$

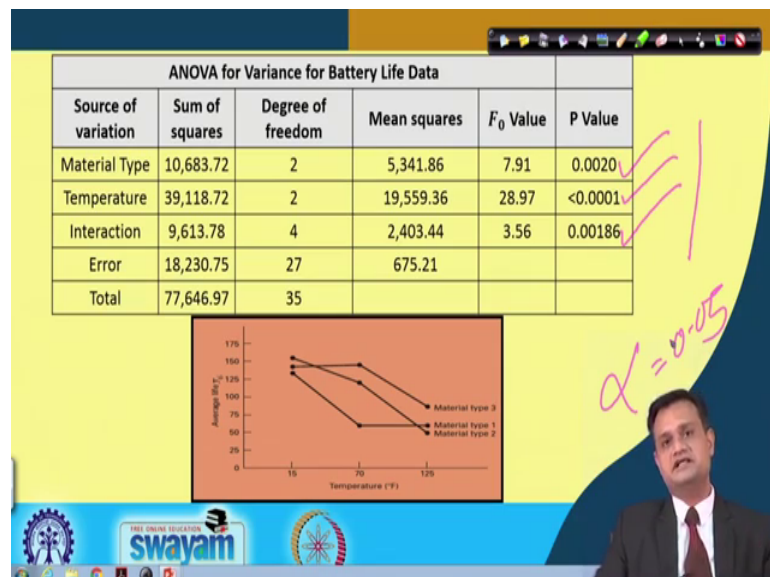
and

$$SS_E = SS_T - SS_{Material} - SS_{Temperature} - SS_{Interaction}$$

$$SS_E = 77,646.97 - 10,683.72 - 39,118.72 - 9613.78 = 18,230.75$$

The slide also features the Swamyam logo and a video feed of the presenter.

(Refer Slide Time: 28:03)



Slide content showing ANOVA results and a line graph:

ANOVA for Variance for Battery Life Data					
Source of variation	Sum of squares	Degree of freedom	Mean squares	F ₀ Value	P Value
Material Type	10,683.72	2	5,341.86	7.91	0.0020
Temperature	39,118.72	2	19,559.36	28.97	<0.0001
Interaction	9,613.78	4	2,403.44	3.56	0.00186
Error	18,230.75	27	675.21		
Total	77,646.97	35			

Below the table is a line graph showing Average life (h) versus Temperature (°F) for three material types. The graph shows that average life decreases as temperature increases for all material types. Material type 3 has the highest average life, followed by material type 1, and then material type 2.

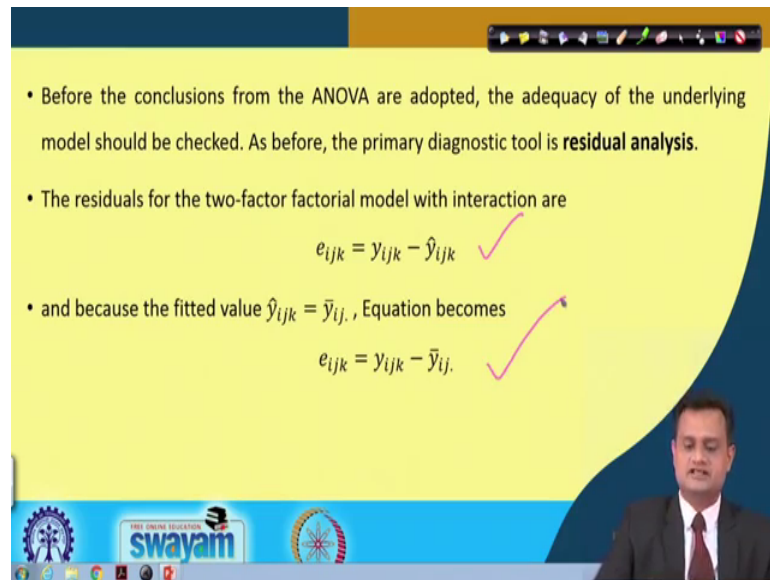
Handwritten notes on the slide include "α = 0.05" and checkmarks next to the p-values in the ANOVA table.

The slide also features the Swamyam logo and a video feed of the presenter.

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)



• Before the conclusions from the ANOVA are adopted, the adequacy of the underlying model should be checked. As before, the primary diagnostic tool is **residual analysis**.

• The residuals for the two-factor factorial model with interaction are

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} \quad \checkmark$$

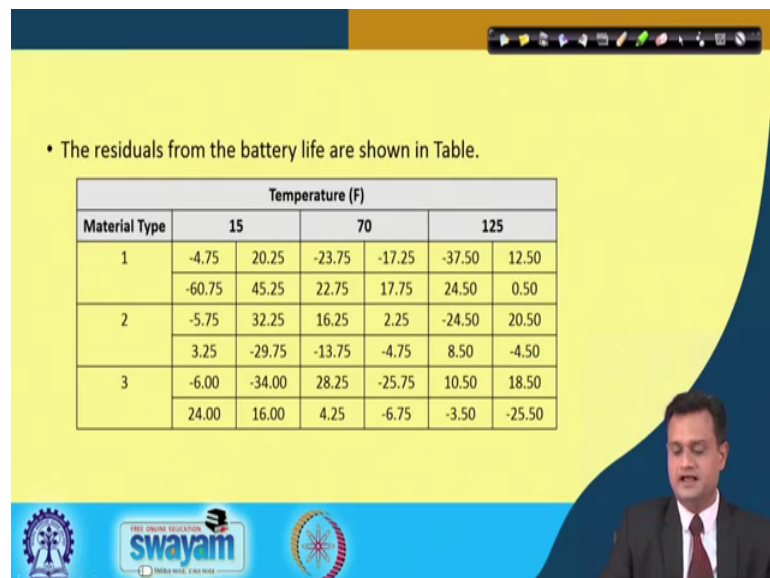
• and because the fitted value $\hat{y}_{ijk} = \bar{y}_{ij.}$, Equation becomes

$$e_{ijk} = y_{ijk} - \bar{y}_{ij.} \quad \checkmark$$

The slide features a yellow background with a blue header and footer. The footer includes the Swayam logo and the text 'FREE ONLINE EDUCATION' and 'INDIA'S MOOC PLATFORM'. A small video inset of a man in a suit is visible in the bottom right corner.

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

(Refer Slide Time: 29:04)



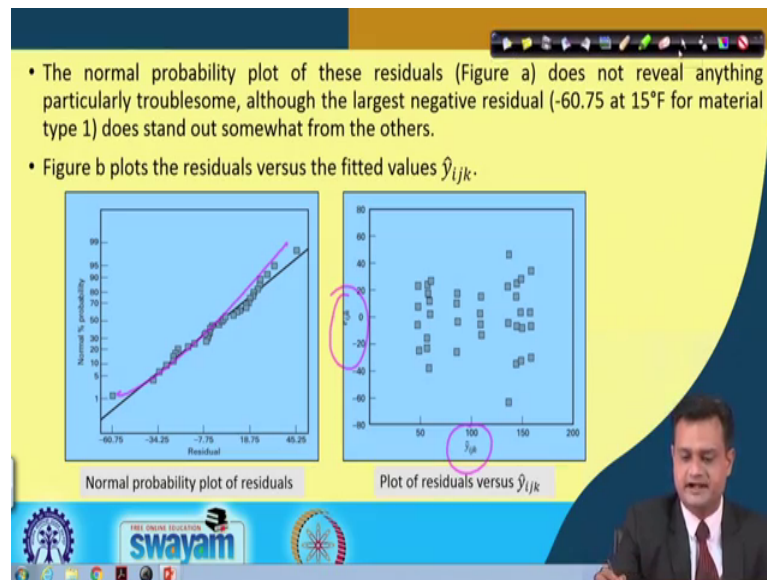
• The residuals from the battery life are shown in Table.

Material Type	Temperature (F)					
	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

The slide features a yellow background with a blue header and footer. The footer includes the Swayam logo and the text 'FREE ONLINE EDUCATION' and 'INDIA'S MOOC PLATFORM'. A small video inset of a man in a suit is visible in the bottom right corner.

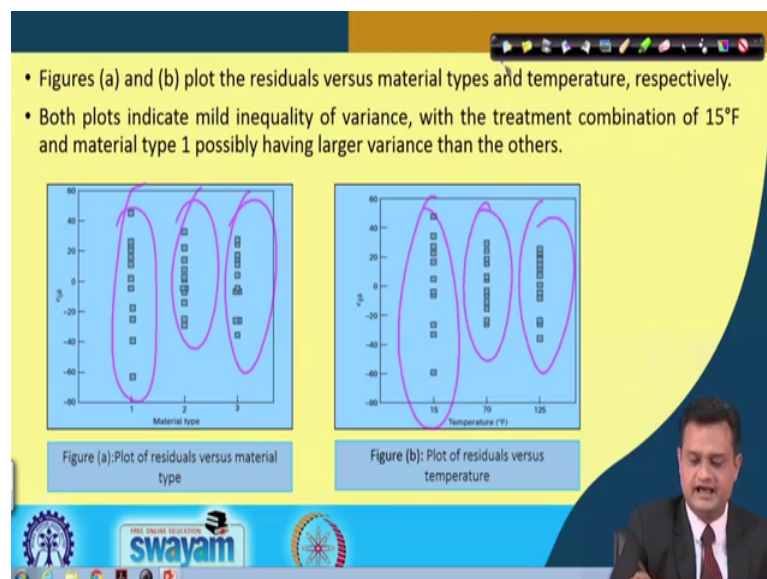
And you can find the residual for each particular material by subtracting it from the grand mean.

(Refer Slide Time: 29:12)



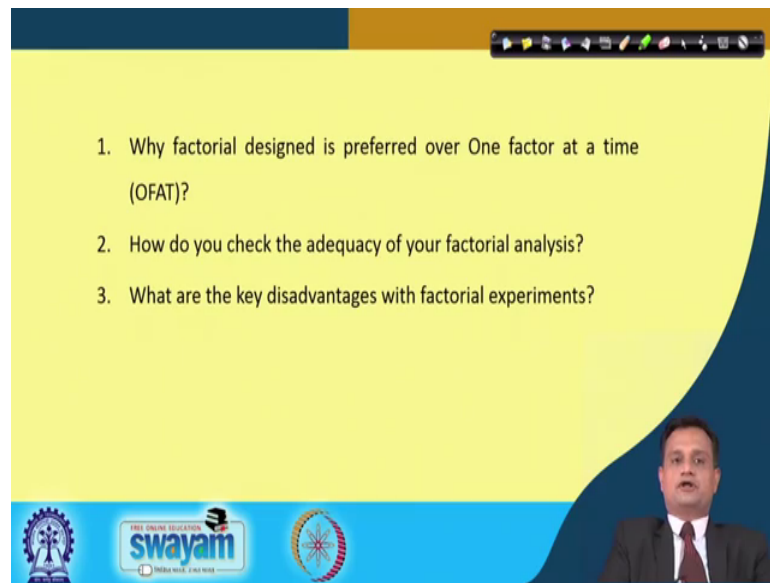
And you can plot it. So, more or less this is going through the line and you have the normality assumption varied say valid and this is when you plot, you can see the scattered net. So, there is an independence which is also verified.

(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)



1. Why factorial designed is preferred over One factor at a time (OFAT)?

2. How do you check the adequacy of your factorial analysis?

3. What are the key disadvantages with factorial experiments?

swayam
INDIA WIDE, TIME WIDE

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.

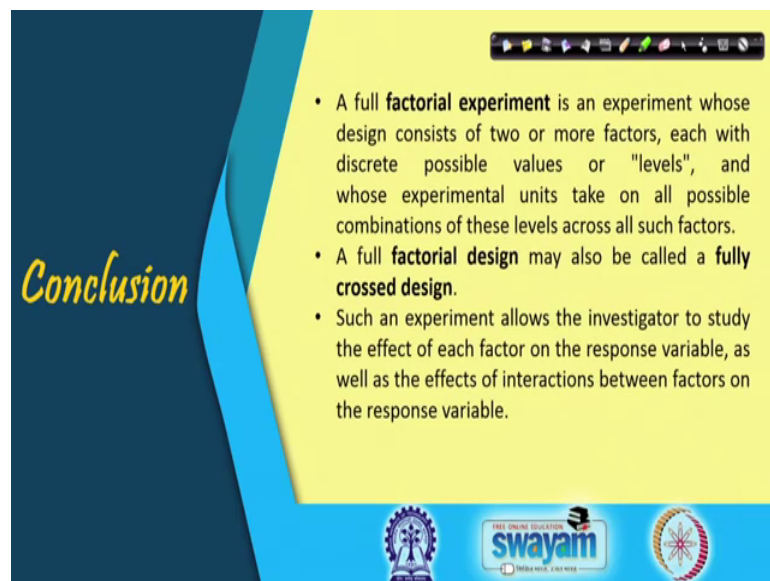
Simultaneously consider the interaction effect and basically it is all about revealing more information about the factors and their interaction effects, so that I can design my processes product with greater accuracy and that can help me to have robust product and design when it will be delivered to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