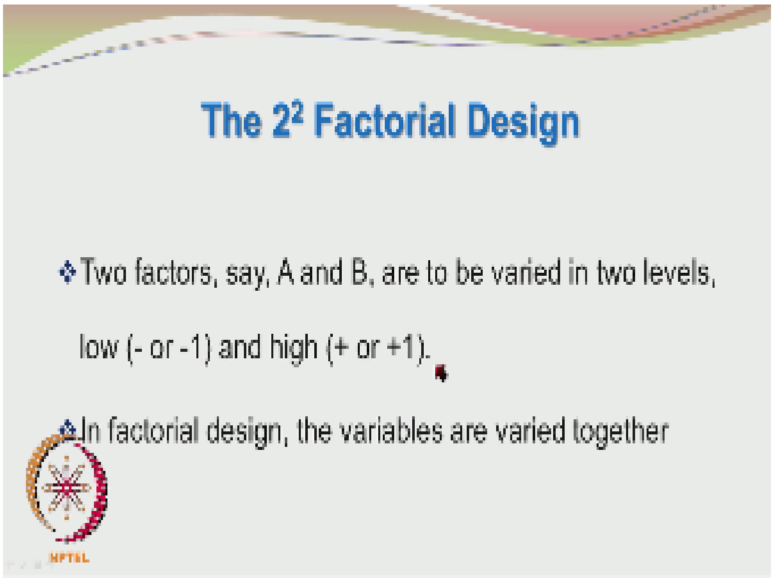
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

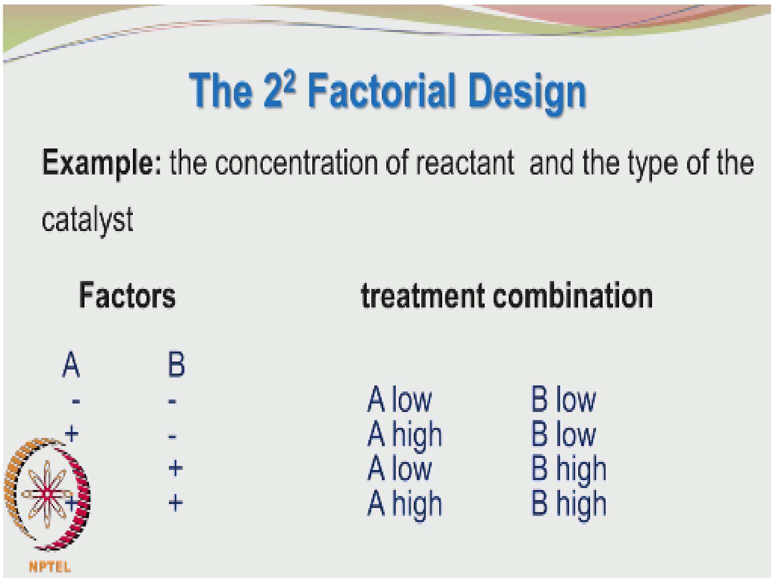
Lecture – 31
Factorial Design of Experiments - Part B: 22 Factorial Design

(Refer Slide Time: 00:19)



Welcome back to the second half of this lecture, we are looking at the factorial design of experiments. First, we will start with a simpler design involving 2 factors at 2 levels; the superscript 2 refers to the number of factors please remember that, so we are going to have 2 power 2 design that means only 4 experiments to see the effects of 2 factors. So, let us call the levels of these factors as - or -1 and the + or +1.

(Refer Slide Time: 00:53)

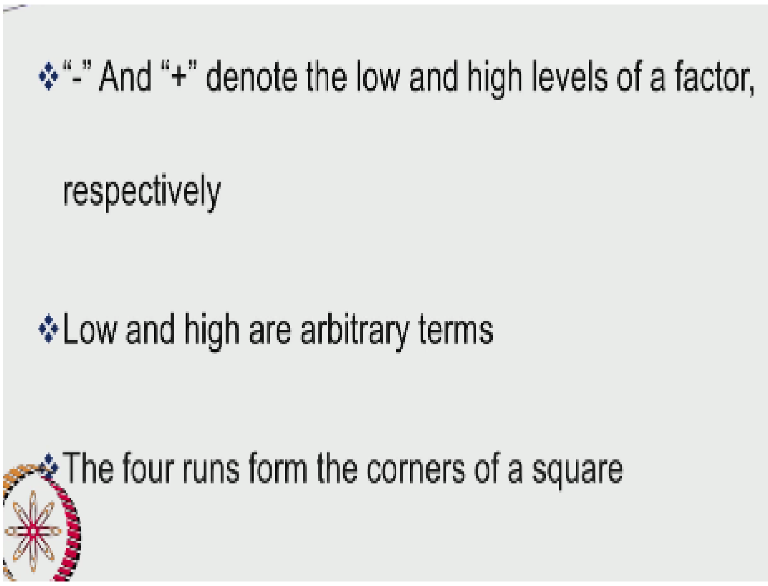


Lower level means - or -1, higher level means + or +1, let us take an example the 2 factors may be concentration of the reactant and the type of catalyst used in a reactor, we are looking at the percentage conversion or the yield, so the 2 factors are shown here; -, - means that both the factors are at their lower levels, + - means only A is at higher level, B is at a lower level, - + means A is at a lower level, B is at a higher level, + + means both of them are at their high levels.

So, this is the coded form of these levels of these factors and the description is given here. Coding is very important in factorial design, you have to convert the actual data given in terms of temperatures like 30 degrees, 60 degrees and so on and the catalyst type A, catalyst type B and so on, you have to convert them into coded form. For example, A and B types of catalyst may be represented arbitrarily as -1 for A and +1 for B.

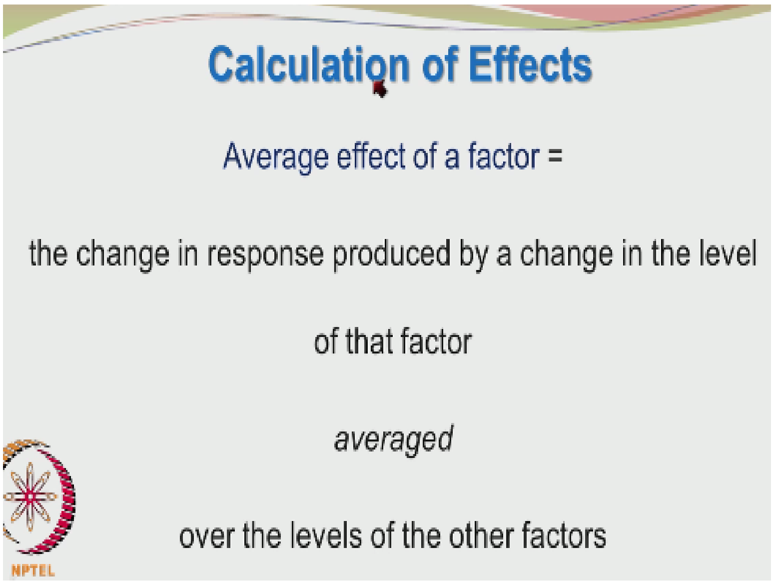
It does not mean that the superior catalyst should be given +1 and the inferior catalyst should be given -1, we do not know which of the 2 catalysts is better, so arbitrarily you can give A as -1 and B as +1 and logically, you can give 30 degree centigrade as -1 and 60 degree centigrade as +1. If you want to give 60 degree centigrade as -1 and 30 degree centigrade as +1 also, it is perfectly fine.

(Refer Slide Time: 02:47)



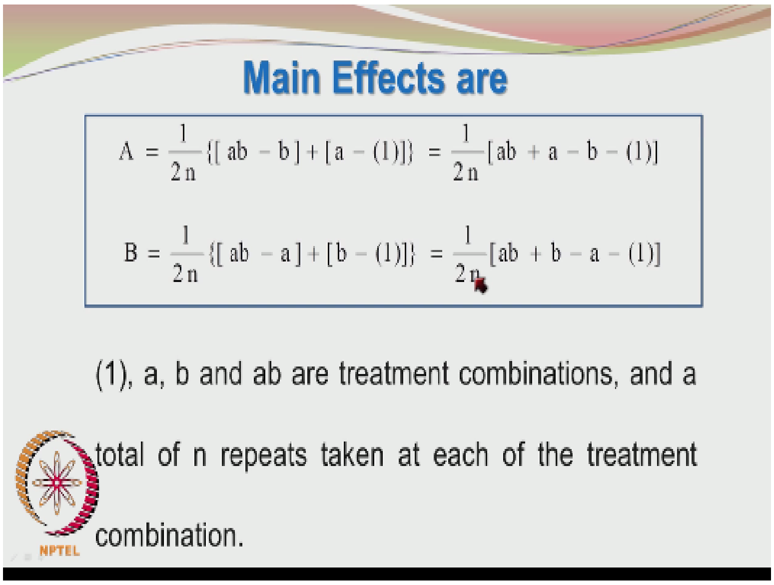
But it is sort of counterintuitive and so we may go for alphabetical ordering for -1 and +1 and numerical ascending order for -1, +1 for numeric factors. So, let us move on that 4 runs, we have considered will form the corners of a square.

(Refer Slide Time: 02:55)



And when we want to calculate the effects, the average effect of a factor this is quite important. The change in response produced by a change in the level of that factor averaged over the levels of other factors, so we want to see the effect of factor A, obviously we have to see the effect of A, when it goes from its lower level to its higher level but in factorial design of experiments, you are not keeping all other variables at constant values, they are also getting varied.

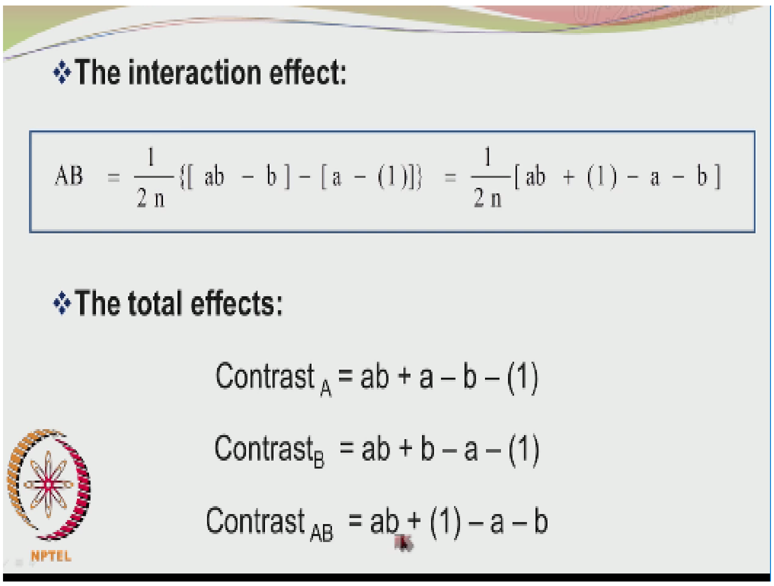
(Refer Slide Time: 03:54)



So, when A is changing from a lower level to higher level, other factors may also be changing and when the other factors are changing also, A may be moving from - level to + level, so the way to calculate the effects is quite simple however, we have to take the average of all those cases, where A moved from lower level - to higher level +. So, in the case of a factorial design, we have ab – b.

Here, you can see that a, was moving from lower level to the upper level, b means, a set level of - and ab means, a is it a level of +. Here, this is one combination, another combination is when all the factors were at the lower level, we had 1 and then when a, alone was at a higher level, we had a. So, this is another case, where a or factor a moved from its lower level to a higher level, another case where factor a moved from its lower level to a higher level.

(Refer Slide Time: 04:57)



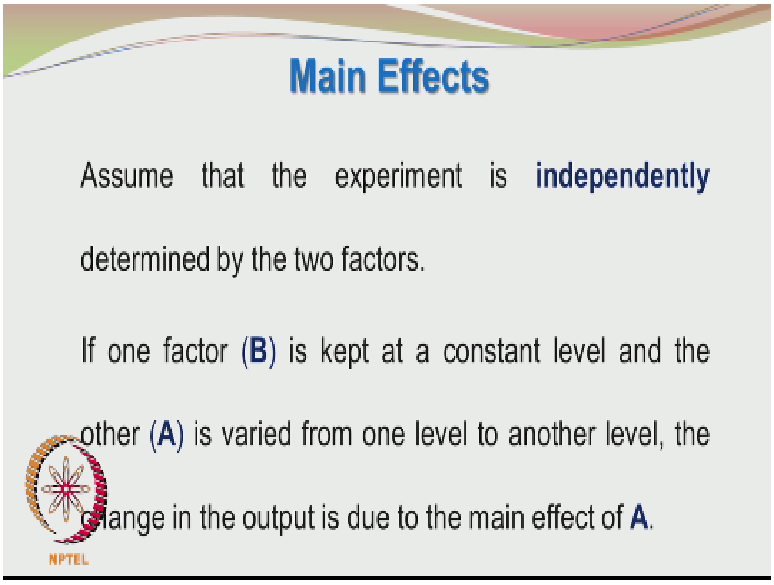
So, these 2 cases we are averaging and then we are multiplying by n to take into account the number of repeats. Similarly, we do for b, so 1; a, b and ab are treatment combinations and a total of n repeats taken at each of the treatment combination, so we also have to account for the number of repeats, when we are doing the averaging. The interaction effect is calculated in terms of ab - b - of a – 1, this is again quite simple.

What does ab interaction tell us, what is the effect of factor a at 2 different levels of factor b? That is all ab interaction is all about okay, so the effect of factor a at higher level of factor b is given by this combination. Here as well as her, b is at a higher level and so you have ab - b giving the change in a, at a higher level of factor B, so we get this difference and then we want to see the change in factor A at a lower level of factor b and that is given by a – 1.

The difference between these 2 gives the interaction, if the change in factor A is 10 units at higher level of b and the change in factor A is again 10 units at lower level of factor B, we get 10 – 10, we get 0; we get ab interaction as 0. So, there is no interaction between factors A and B but at higher level of b, we get a affect as 10 units and at lower level of factor B, we get the a effect as 15 units, then it becomes 10 – 15, we get -5.

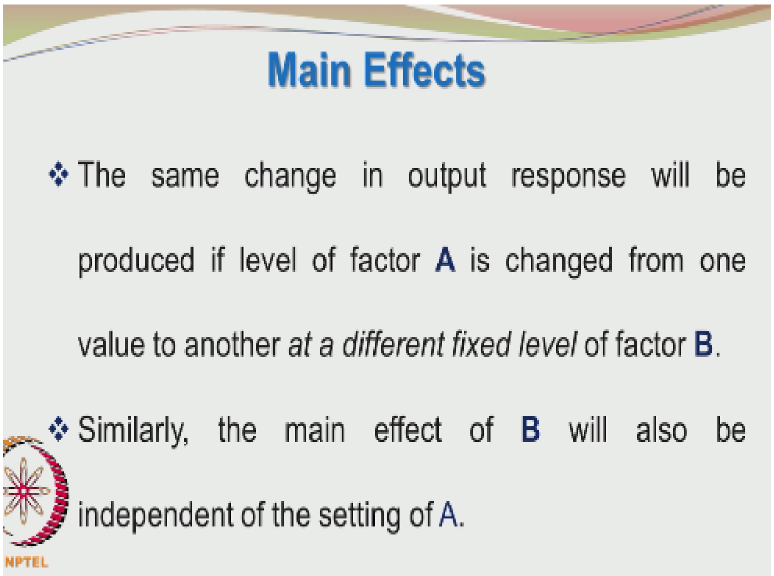
That means, there is a significant interaction between factors a and b. If you remember this concept, this is sufficient okay. When we go for more number of factors, we get more and more number of terms and it may be a bit difficult to relate to them but the same concept applies in all of them and we call the term inside the brackets or parentheses as contrasts. So, ab + a – b - 1 is a contrast.

(Refer Slide Time: 07:29)



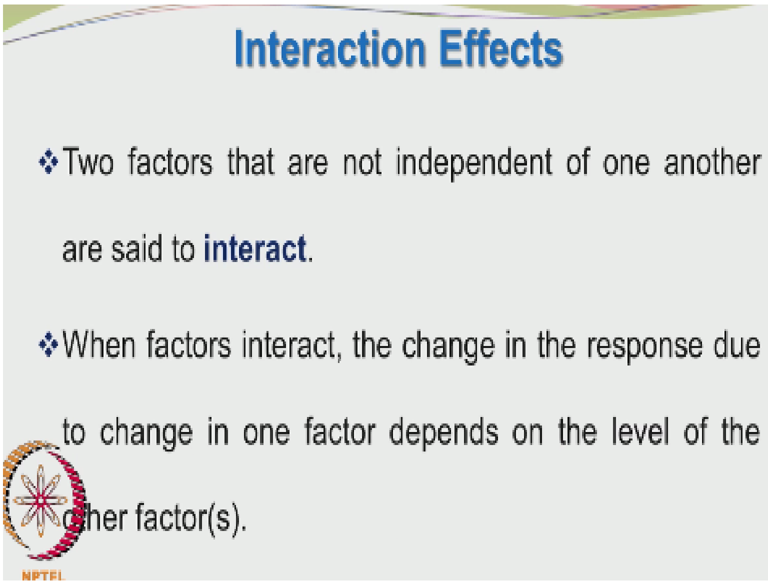
Similarly, ab, we have ab + 1 - a - b is also a contrast, so for contrast for A is ab + a – b – 1, contrast for B is ab + b - a – 1, contrast for AB is ab + 1 - a - b and that is what I said, so we want to calculate the effects of factor A and if one factor B is kept at the constant level and the other is varied from one level to another level, the change in the output is due to the main effect of A.

(Refer Slide Time: 07:51)



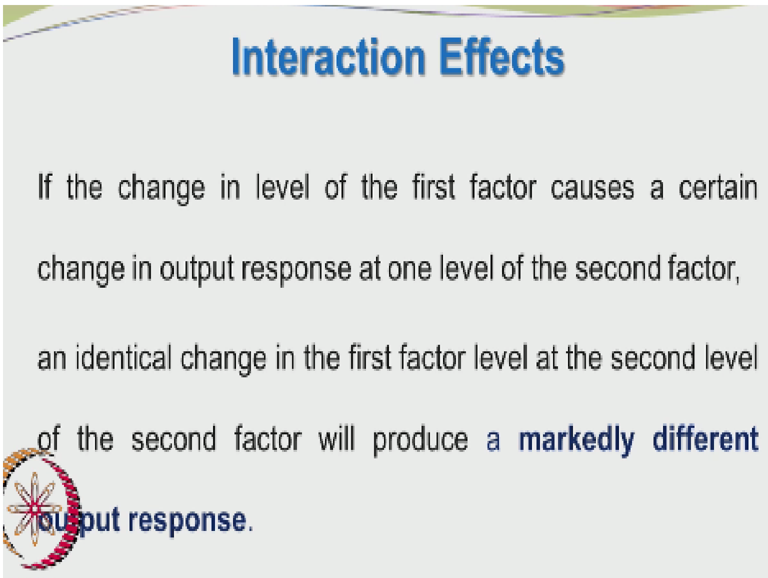
The same change in output response will be produced if level of factor A is changed from one value to another level at a different fixed level of factor B. So, similarly the main effect of B will also be independent of the setting of A, this is when the 2 factors do not interact.

(Refer Slide Time: 08:08)



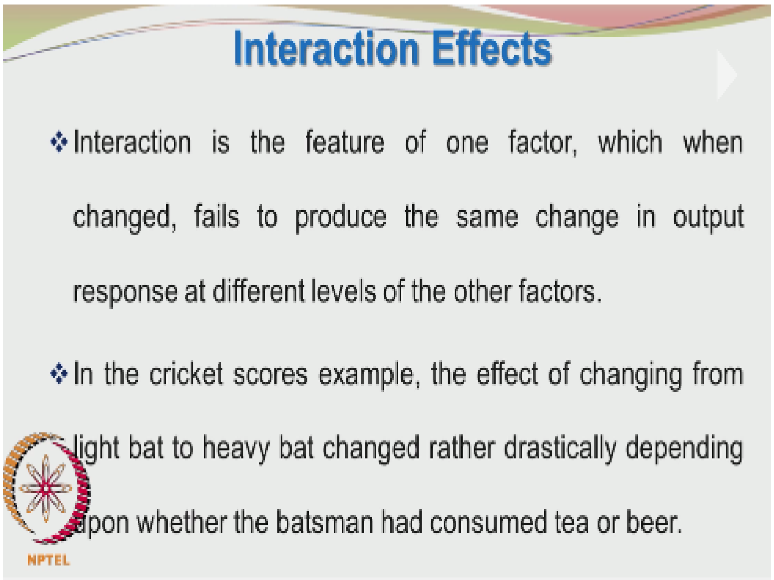
However, when the 2 factors that are not independent of each other are said to interact, when factors interact the change in response due to change in one factor depends on the level of the other factors, this is again quite self-evident.

(Refer Slide Time: 08:22)



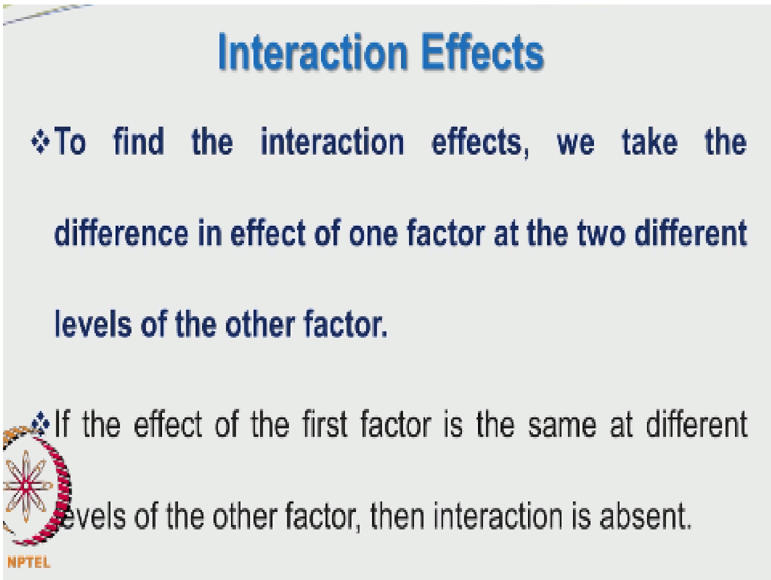
So, we are now still talking about interaction effects, if the change in level of the first factor causes a certain change in the output response at one level of the second factor, an identical change in the first factor level at the second level of the second factor will produce a markedly different output response, when there is interaction.

(Refer Slide Time: 08:45)



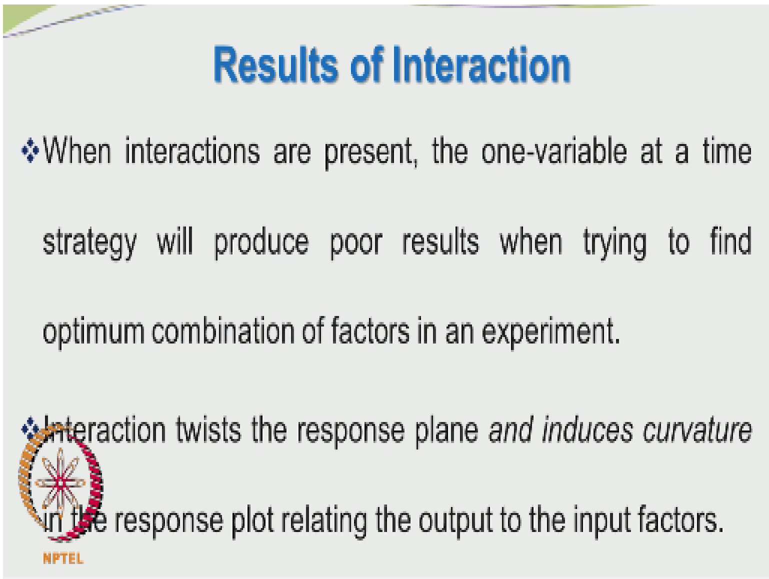
What is interaction? Interaction is the feature of one factor, which one changed, fails to produce the same change in output response at different levels of the other factors. In the cricket scores example, the effect of changing from light bat to heavy bat changed rather drastically depending upon whether the batsman had consumed tea or beer. I am going a bit fast here because the concepts I have already explained the reference to the 2 power 2 design.

(Refer Slide Time: 09:16)



It is just being said in different ways here, so to find the interaction effects, we take the difference in the effect of one factor at the 2 different levels of the other factor. If the effect of the first factor is the same at different levels of the other factor, then interaction is absent.

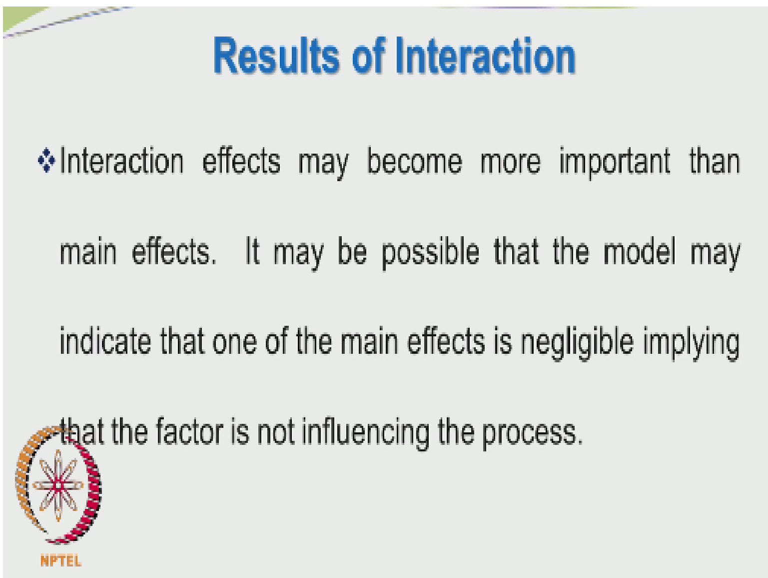
(Refer Slide Time: 09:33)



When interactions are present, the one variable at a time strategy will produce poor results, when trying to find the optimum combination of factors in an experiment. Interaction twists the response plane and induces curvature in the response plot relating the output to the input parameters. The simple summary here is if you have interaction between the factors, the simple strategy of one variable at a time experimentation will fail to find the true optimum conditions.

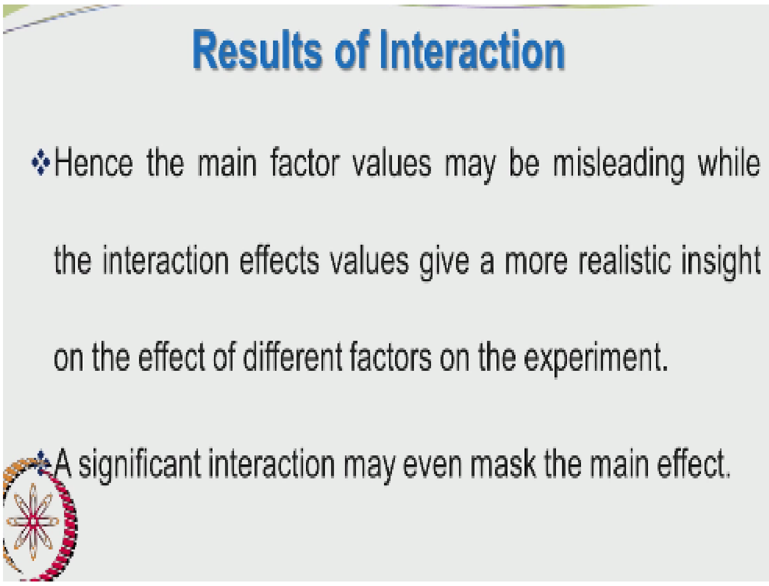
So, you have to resort to design of experiments to not only identify the interaction effects between the involved factors but also proceed using the design of experiment strategy to find the optimum combination accurately otherwise, you will be far away from the optimum using the one variable at a time approach and that is not good for the process okay. It will be suboptimal leading to loss in time, production, energy, money, labour and so on.

(Refer Slide Time: 10:41)



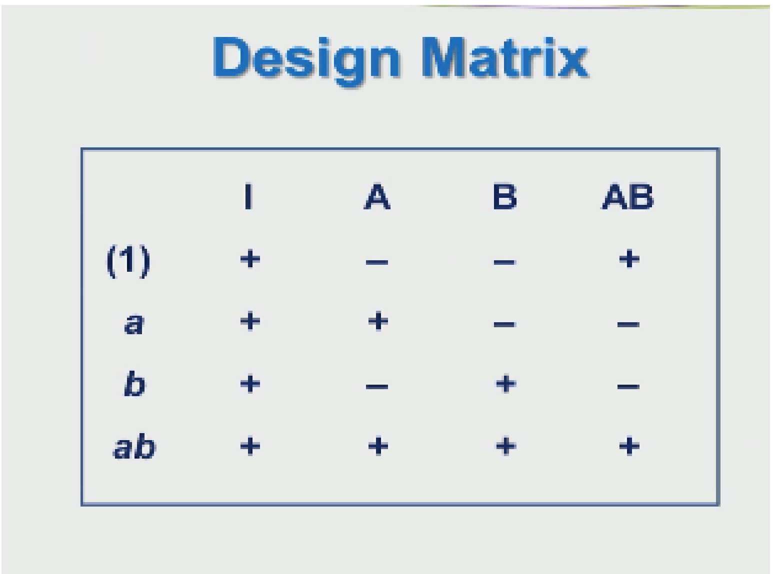
Actually, interaction effects may be more important than the main effects and even there may be certain cases, I have seen one example, where the main effect is unimportant. Suppose, you are having 2 factors A and B, one of the factors is unimportant. A is for example unimportant but when you are looking at the interaction between A and B; AB may be having a significant role to play.

(Refer Slide Time: 11:31)



So, you cannot conclude or generalize that when one main factor is absent, all the interactions involving that factor would be insignificant. Even though, a main factor may be insignificant when it is acting together with other factors, it may have a significant interaction. So, it is always better to look at the interaction effects rather than at the main effects, only when tests have concluded or tests have conclusively shown that the interaction effects are insignificant.

(Refer Slide Time: 12:05)



Design Matrix

	I	A	B	AB
(1)	+	–	–	+
a	+	+	–	–
b	+	–	+	–
ab	+	+	+	+

Then, you pay attention to the main factors and look at the contribution of the different main factors to the overall outcome of the process and very importantly, a significant interaction may even mask the main effect. Now, let us look at this design matrix, this is very interesting and it also helps us to find the effect of different factors and the interactions in a very simple manner. So, this particular design matrix is very important.

Here, we have setting 1 and a, b, and ab, this I is the identity column, if you want to call it that way and it is having all entries as positive, then you have A, when you look at 1, at 1, both A and B are at their lower settings, so you put minus here and you also put minus here. AB would be product of A and B and that would be minus into minus which is plus. When you look at small a, it means that only factor A is at the higher setting.

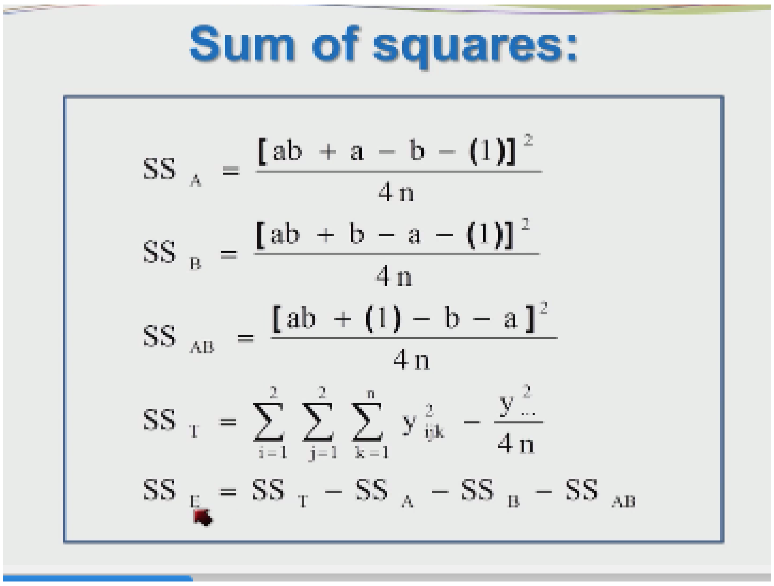
And all other factors are at the lower setting, so you get a, you put plus for A only, for B it is minus and so AB would be minus, b is the other way, A is at a lower level minus, B is at a higher level plus, AB is still negative and ab, both of them are positive and so, AB is also positive and this is very interesting and when you add up all these minuses and pluses, you will get 0. The number of minuses will compensate for the number of pluses.

So, any of these columns will have 0, when you add up the minuses and pluses that is one important thing and when you look at A, you are having -1 + a – b + ab, -1 + a – b + ab, so A and ab are positive 1 and b are negative, you go to the contrast for A, we saw that a while back, if you look at the contrast for A, you will find a and ab are positive and b and 1 are negative, this is exactly what we saw from the design matrix.

So, you can use that to calculate the effect of A; ab + a – b – 1, then you divide it by 2 and then you multiply by the number of repeats that will give you the effect of main factor A. Similarly, you can do the same approach or use the same approach to find the main factor B, which is given by ab + b – a – 1. What has happened is; we have switched a instead of b and so you have +b and then you have –a.

So, you can use the design matrix itself to identify the contrast set and then you divide by the number of repeats and also average them out, so you are having two sets ab - a and b -1, so you can put it as 2 here. So, this is a very simple way, you can do it not only for the main factors but you can also do it for the interaction. For interaction, we see it is ab + 1, the extremes are positive and the intermediate ones are negative, both a and b are negative.

(Refer Slide Time: 15:38)

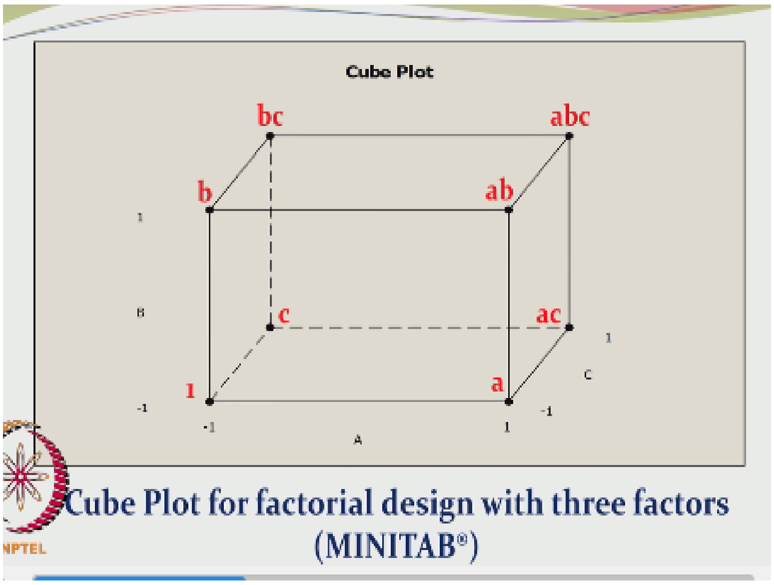


Let us go and look at the design matrix, so the extremes are positive and a and b are both negative. So, using this design matrix, we can calculate the contrast and hence the factors pretty easily. Now, let us look at the sum of squares not only we have to look at the main effects and the interactions, we also have to see which one of them is more important than the others and for that, we need to do the analysis of variance and for that we have to calculate the sum of squares.

Calculation of the sum of squares is also pretty easy, we again use the contrasts and square the contrast and divided by 4 * number of repeats. Why we have 4 here? There is a bit of theory involved using the approach of contrasts and how to calculate the sum of squares from them. I am not getting into that it will take us away from our area of focus; there is a good discussion on the use of contrasts in Montgomery's design and analysis of experiments book.

You may please refer to that to get an idea on why we should put 4 here and when you go for higher order factorials, why we should put the certain coefficient; we will come to that when we look at 2 power 3 designs. Similarly, for sum of squares of B, we take the contrast for B, square it and then divided by 4n, again for sum of squares of AB; we take the contrast for AB, square it and then divide it by 4n.

(Refer Slide Time: 17:21)



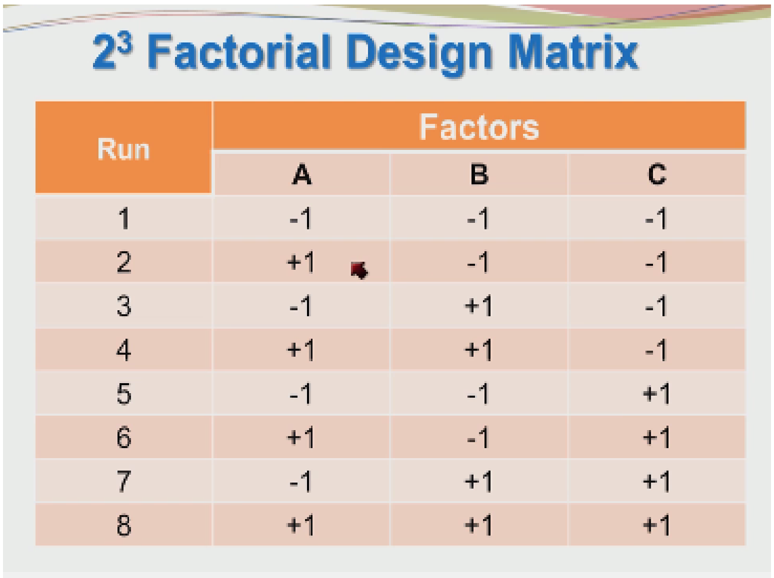
And total sum of squares is given by the shortcut formula and the sum of squares of error is given by the difference between the total sum of squares and the sum of squares contribution from A, B and AB and this gives us the sum of squares of the error. So, now let us look at factorial design involving 3 factors, since there are 3 factors, the design would become 2 power 3, where 2 is the number of levels; -1, +1, lower level or upper level.

And then we are having 3 factors, so we are having 2 power 3 design that means it will involve 8 runs and these 8 runs may be represented as the corners of a cube and now it should be easy for you to identify the various components; 1 means, all a, b and c are at their lower levels, you can take this as a 2 dimensional representation of a 2 power 2 design involving factors a and c. So, the bottom base would be 2 power 2 design involving factors a and c.

So, this means only a is at a higher level, this means c is at a higher level and this ac means both a as well as c factors are at their higher settings and so this forms the corners of a square and this particular 2 dimensional representation represents, the 2 power 2 design involving factors b and c. Similarly, you can see the significance of the various corners. For example, at this ab, it means that factors A and B are both at their higher levels go along this direction.

We will see factor A is at a higher level go along this direction, you see factor B is at a higher level, so both of them are at higher levels, so ab combination we have, which means also that factor C is at a lower level. Since, you are going for c in this direction, factor C would be at a lower level and so you do not find C but when you have this particular point, all of them are at their higher levels and so you have the notation abc.

(Refer Slide Time: 19:27)



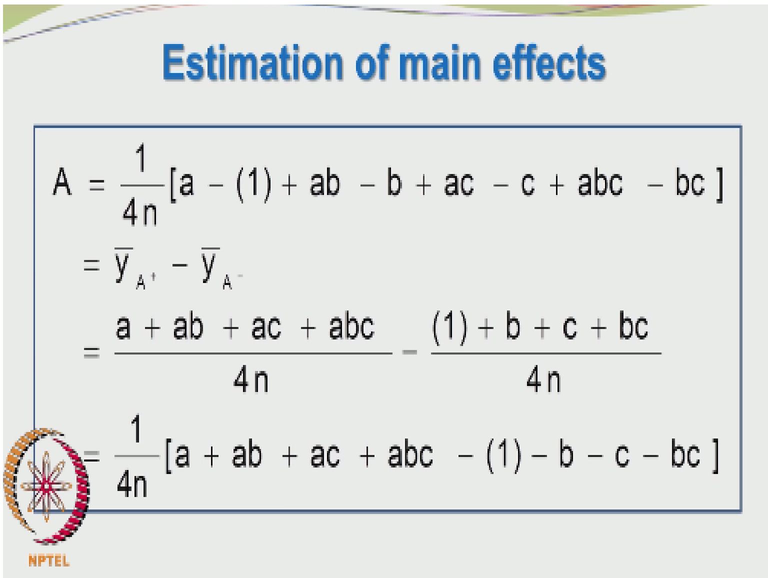
2^3 Factorial Design Matrix

Run	Factors		
	A	B	C
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

It is very simple, if you have understood the 2 power 2 design, there will be very little difficulty to understand this particular concept. Now, let us look at the similar 2 power 3 factorial design matrix, this is very interesting, you can see that ABC are the lower levels here and then I am changing only A, so +1, -1, -1 and very importantly, I am not changing it only one variable or one factor time.

When I am going from run 2 to run 3, you may see that I have changed A from +1 to -1 and B from -1 to +1, keeping C level at -1, so you can see that I am changing the factors simultaneously. Similarly, when I am going from 4 to 5 or even from 3 to 4, I have again changed A from -1 to +1 and B and C are at their lower levels but when I am going from 4 to 5, I have changed both A and B from +1 to -1.

(Refer Slide Time: 20:52)

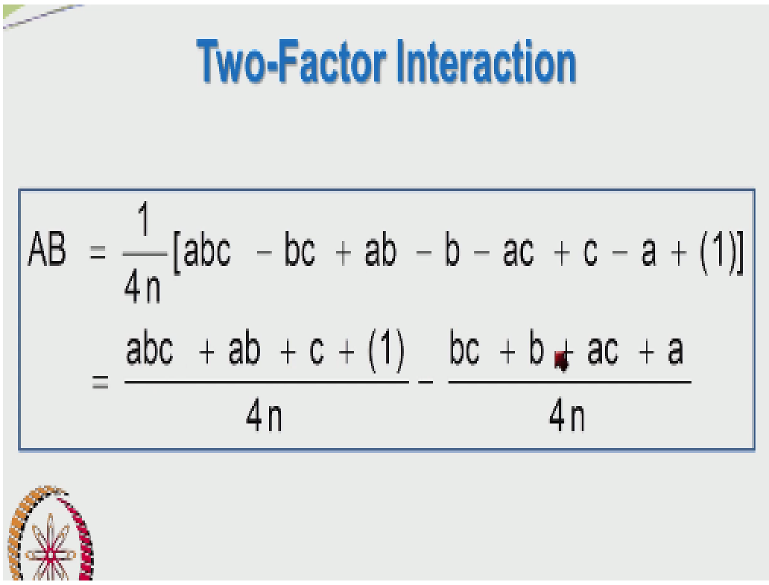


In fact, I have even changed C from -1 to +1, so from levels 4 to level 5, I have changed all the factors simultaneously, this is one important feature of the 2 power n fact or 2 power K factorial design, where K is the number of factors. In one variable at a time, you are going to change only one variable from one level to another level keeping the other variables at the fixed values. Moving on, we can see that now the number of factors has increased from 2 to 3.

And so the elements are looking more cluttered, it is slightly more difficult to explain the difference between different levels, when A changes, so you can see that there are 8 elements; 1,2 , 3, 4, 5, 6, 7, 8 and the same concept applies. What is the average value of A at higher setting of A - the average value of A at the lower setting of A. So, let me come again. What is the average response at the higher level of A and what is the average response at the lower level of A?

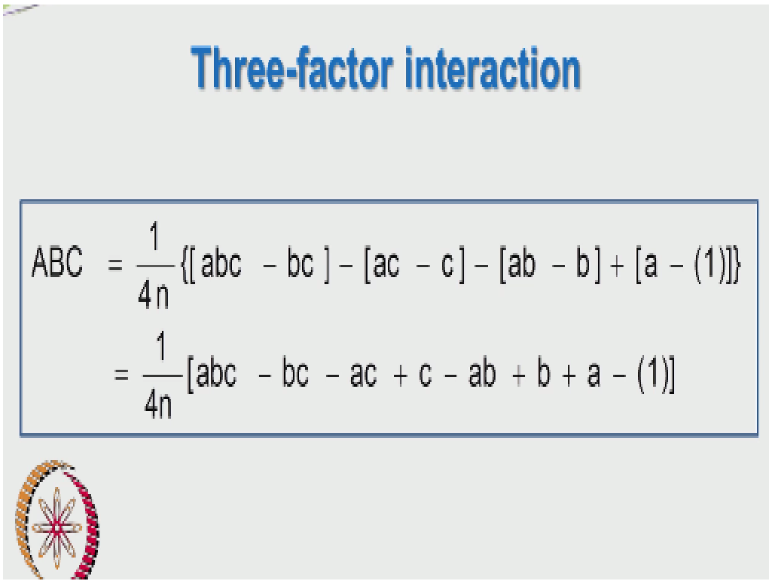
And I am then taking the difference of these 2 to get the effect of A and so you may think, oh my god, I have to remember such large formulae for the quizzes and exams, you do not have to, you can see that everything involving a is positive; a, ab, ac, abc is positive and everything not involving a is negative; 1, b, c and bc. You please go back to the design matrix and okay, we have to construct the design matrix for this case, we will do so in a minute.

(Refer Slide Time: 22:22)



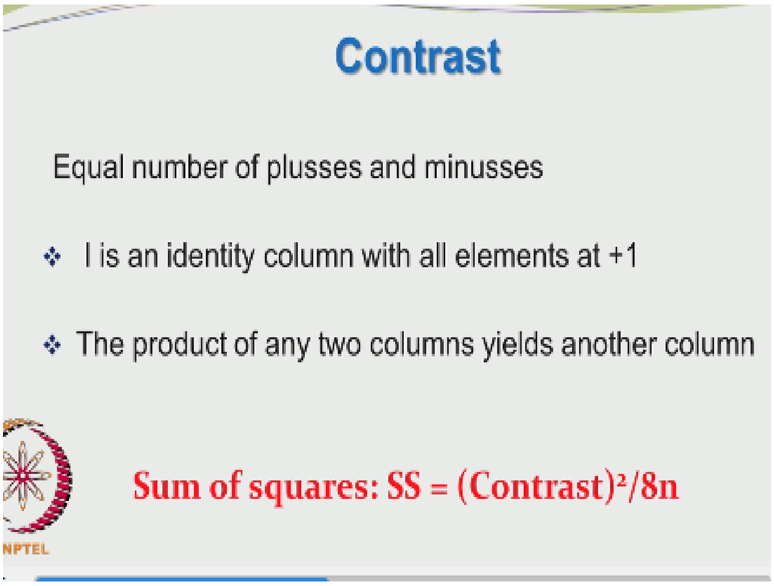
Similarly, for AB, you can see that these are positive and these are negative, again you do not have to break your head on trying to remember such big formulae and you may start wondering what should I do if I have 2 power 4 design, which will involve 16 elements and that would become very cumbersome to remember.

(Refer Slide Time: 22:44)



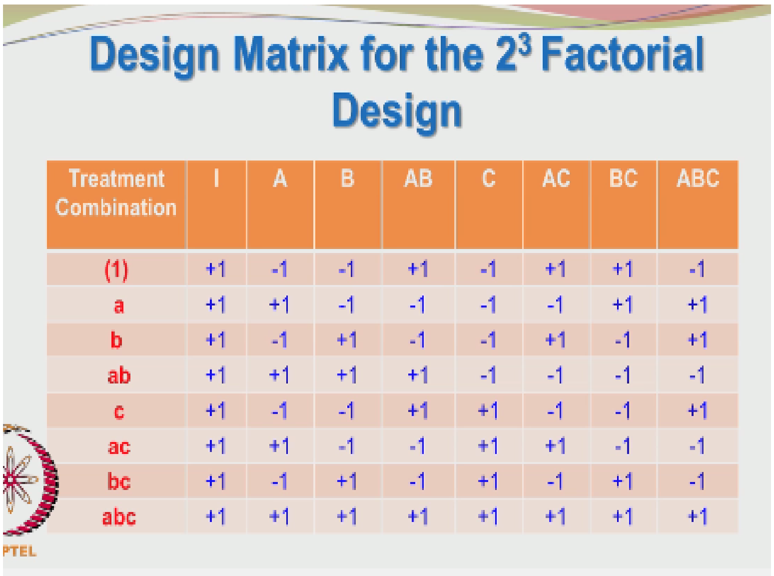
So, you can also have ABC interaction, so this is becoming quite unmanageable.

(Refer Slide Time: 22:51)



But let us look at the design matrix shortly, so you can see the contrast, you will have equal numbers of pluses and minuses, so you have 4 pluses and 4 minuses that is expected and the product of any two columns yield another column, we have already seen this and sum of squares is contrast squared/ 8n, okay, so that is very important. So, how do we set up this contrast?

(Refer Slide Time: 23:27)



Design Matrix for the 2^3 Factorial Design

Treatment Combination	I	A	B	AB	C	AC	BC	ABC
(1)	+1	-1	-1	+1	-1	+1	+1	-1
a	+1	+1	-1	-1	-1	-1	+1	+1
b	+1	-1	+1	-1	-1	+1	-1	+1
ab	+1	+1	+1	+1	-1	-1	-1	-1
c	+1	-1	-1	+1	+1	-1	-1	+1
ac	+1	+1	-1	-1	+1	+1	-1	-1
bc	+1	-1	+1	-1	+1	-1	+1	-1
abc	+1	+1	+1	+1	+1	+1	+1	+1

This is the table or now we are back in business, so you have different treatment combinations, all factors at a lower level; a is at a higher level, b is at a higher level both a and b at a higher level and c at a lower level, only c at a higher level, ac both factors a and c at a higher level but b is at a lower level, bc both factors b and c are at their higher levels and factor a is at a lower level, abc, all factors a, b and c are at their higher levels.

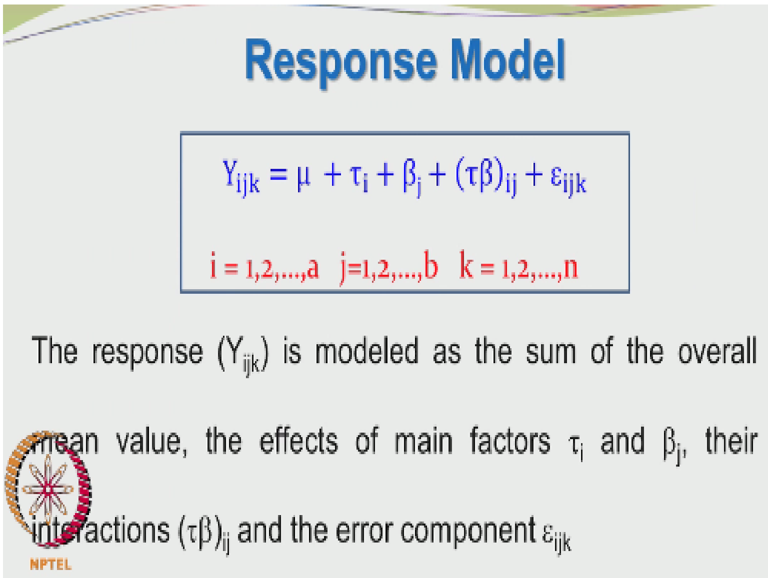
And this is the identity column and that is +1, +1, +1 throughout A, this would be -; a, would be plus, b would be minus, ab would be both a and b at the higher level, so you give +1 here. C; only c is at a higher level, so you give -1 here, ac; a is at a higher level, c is at the higher level, so you give +1 for A; factor A, bc is at a lower level, so we will give -1 here, abc all of them are at a higher level, so you give +1 here.

Using this logic, we can do for all other elements for example, when you have AB, you do not have to worry about what setting AB would be; simply multiply A and B; -1 * -1, you will get +1 for AB. Similarly, for C, at 1, it will be -1 obviously, for b again, it will be negative, so this logic may be extended and applied uniformly and you can get not only the main factors contrasts but also you can get the contrast for the binary interaction and also for the ternary interaction.

For example, when you have ABC here, you multiply A, B and C in this case, it is -1 * -1, which is +1, I mean 2 -1, it will be -1, so you can use either -1 or - or + and +1, so that is not a problem - or -1 refers to the lower setting and + or +1 refers to the higher setting. For example, let me take C; 1 is at a lower level, so C would be -1, only a is at a higher levels, again C will be -1, only b is at a higher level, so C will be again -1.

Ab; both a and b are at their higher levels but C is at a lower level, so again C is at a -1 but from now on, you can see that c is at a higher level throughout and so you will have +1 throughout for C. I hope you have understood this logic, it is very simple, you just try to do the things on your own and very quickly you will get the correct sequences or correct contrast sets for all the main factors as well as the binary and ternary interactions.

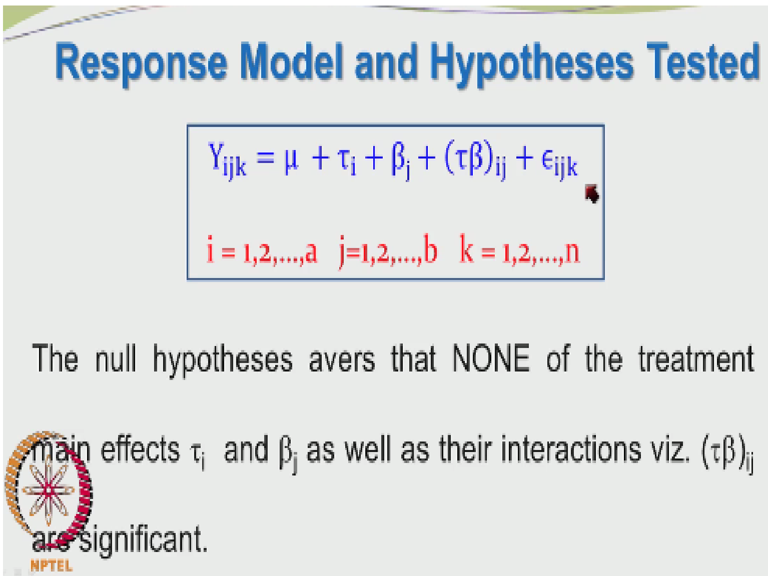
(Refer Slide Time: 26:13)



Now, we are going to look at the response model, we are slowly now moving from identification of the effects to the importance of the different effects or the relative importance of the different effects. So, now let us look at the response model Yijk that is = Mu + tau i + beta j + tau beta ij + epsilon ijk, this is a linear model which is split into the overall mean value Mu.

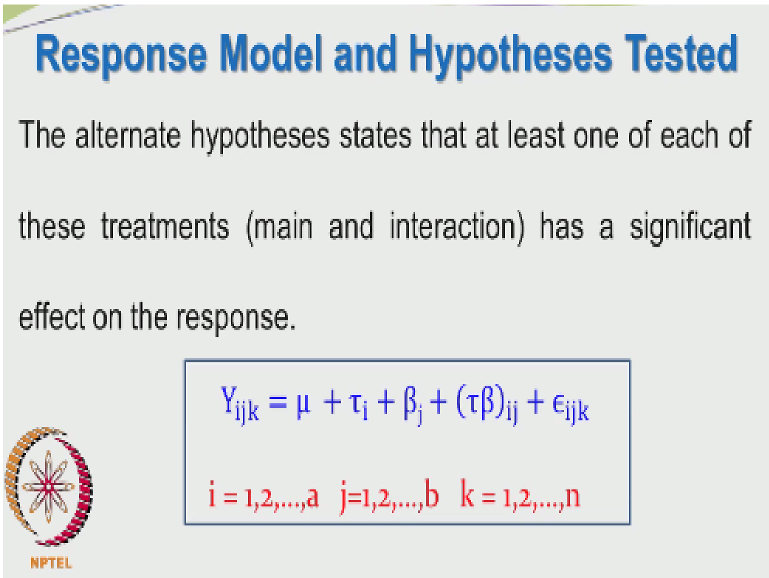
The contribution from factor A, contribution from factor B the interaction contribution tau Beta from the 2 factors and the you have the error component, so you can have a settings of factor A, you do not have to have necessarily 2 settings of factory A; -1 and +1, you can even have 3 settings of factor A and that may be different for the number of settings of factor B and that may be different from the number of repeats.

(Refer Slide Time: 27:35)



Usually, we put small a is = 2, small b, we put it is = 2, n can be any number >1, for meaningful repetition. So, the response Yijk is model as the sum of the overall mean value plus the effects of factors A and factors B plus the effect of the interaction between A and B and the random error component epsilon ijk and what is the null hypothesis? The null hypothesis tells that the response is only based on the overall mean response plus the error component.

(Refer Slide Time: 28:14)

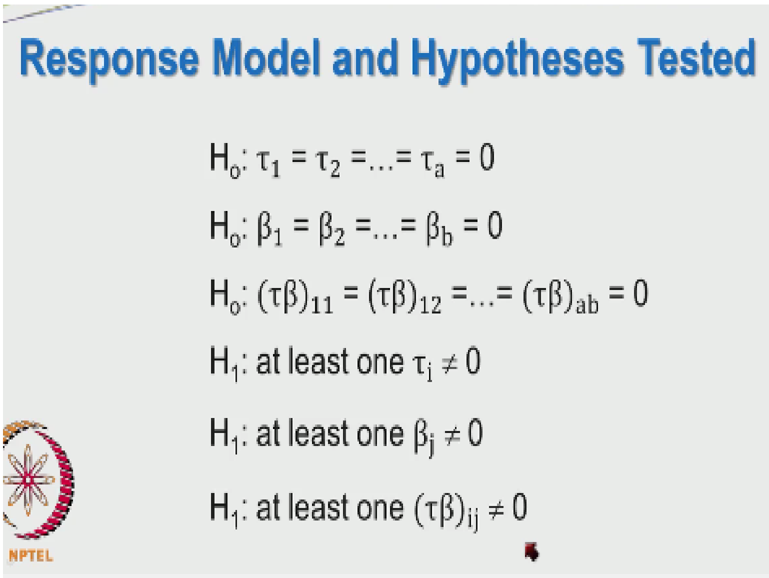


It tells that effect of factor A is 0, effect of factor B is 0 and there is no interaction between the factors, whatever may be the setting of factor A or whatever may be the setting of factor B, their effects are 0 okay. The response is only the overall mean value tempered or altered by the random error component. The alternate hypothesis states that at least one of each of these treatments, main and interactions have a significant effect on the process.

So, you are having this and at least one level of treatment A and one level of treatment B or one level of interaction between the 2 treatments will be effective on the process, so the null hypothesis says that all effects are unequivocally 0, not a single factor has an influence on the response of the process but the alternate hypothesis tells at least one factor is important either tau i is important or beta j is important.

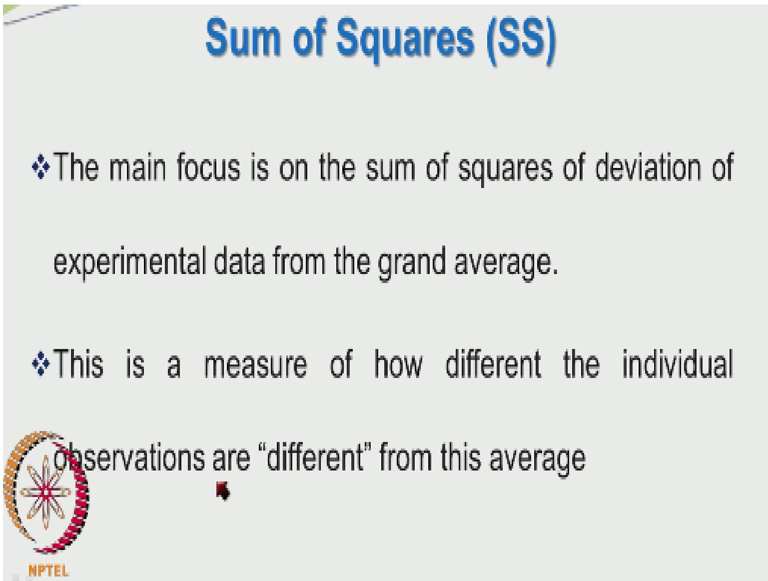
Or at least one interaction between tau i and beta j is important okay, so this is the null and alternate hypothesis. What it means is one of each of these treatments okay, at least one setting involving factor A is having an impact on the process okay. When you are going from one level of factor A to another level of factory A, keeping the second factor B at a constant value, at least we are going from one setting to another setting, in one such cases; one of such cases you are going to have an effect.

(Refer Slide Time: 30:09)



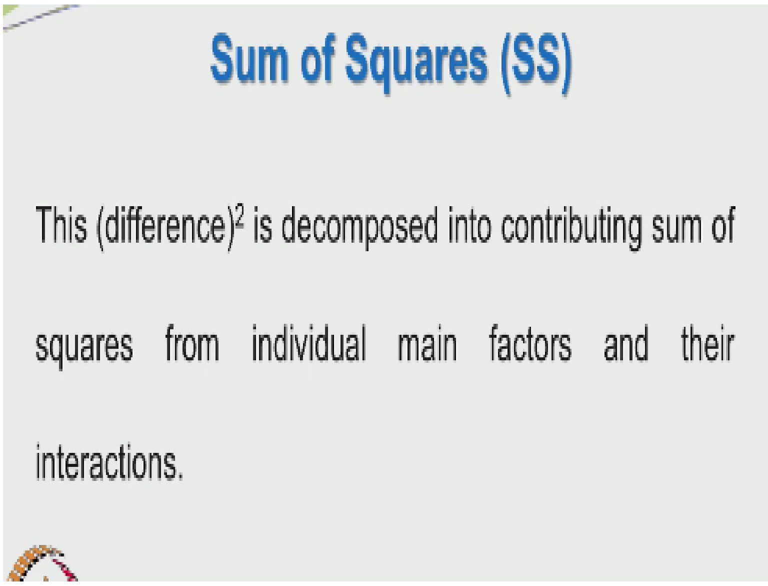
Similarly, when you keep A as constant, when you go from one setting of factor B to another setting of factor B, there is an effect on the process, which is different from the overall mean Mu. So, this is the mathematical statement of all these null and alternate hypotheses, you can say that unequivocally tau1 equals tau2 so on to tau a is = 0 and similarly for the factor B, beta 1 equals beta 2 so on to beta b is = 0.

(Refer Slide Time: 30:43)



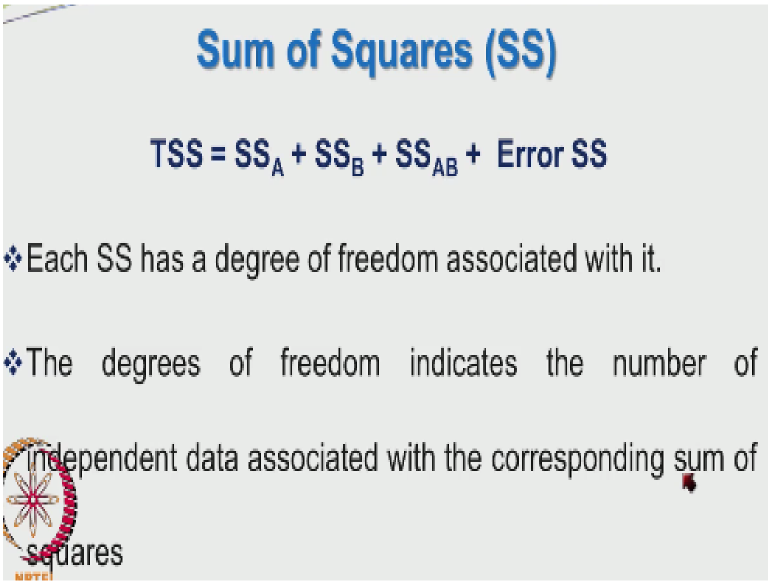
And then, the main interactions are also 0 and then you have the alternate hypothesis saying that at least one tau i is not = 0, at least one beta j is not = 0 and at least one tau beta ij is not = 0. So, how do you get the sum of squares? So, let us now focus on the total sum of squares and it is the sum of the squares of the deviations of the experimental data from the grand average okay. So, this is a measure of how different individual treatments are from this average.

(Refer Slide Time: 31:06)



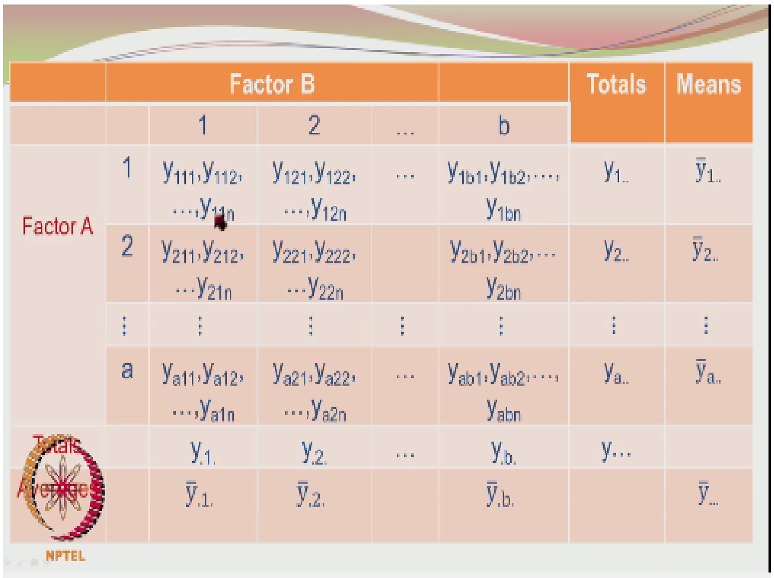
And this total sum of squares may be decomposed or resolved into contributions from factor A, factor B, the interaction between A and B and also the random error component, so once you are able to split these entities, you can compare the contribution from each entity with a total sum of squares.

(Refer Slide Time: 31:37)



So, that is what I have written here, total sum of squares is = SSA + SSB sum of squares of A, sum of squares of B and sum of squares of AB + error sum of squares. Each sum of squares has a degree of freedom associated with it. The degrees of freedom indicate the number of independent data associate with the corresponding sum of squares.

(Refer Slide Time: 32:00)

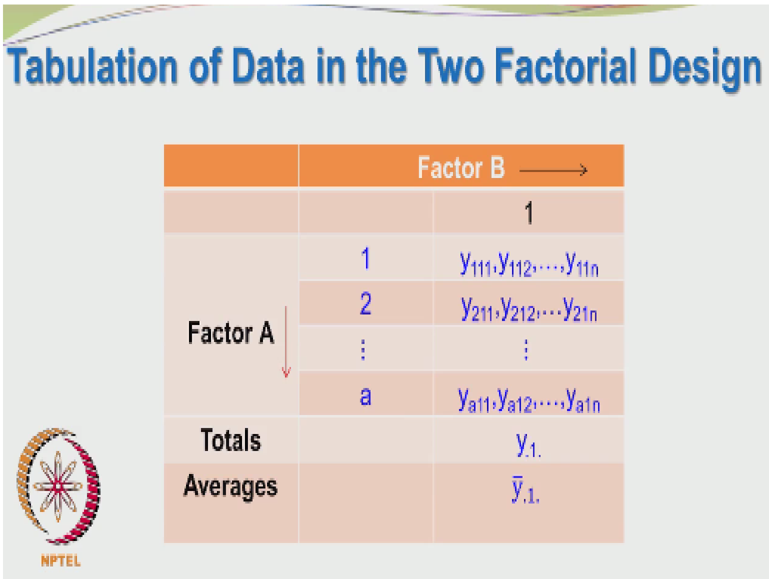


		Factor B				Totals	Means
		1	2	…	b		
Factor A	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$	…	$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$	$y_{1..}$	$\bar{y}_{1..}$
	2	$y_{211}, y_{212}, \ldots y_{21n}$	$y_{221}, y_{222}, \ldots y_{22n}$		$y_{2b1}, y_{2b2}, \ldots y_{2bn}$	$y_{2..}$	$\bar{y}_{2..}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$	…	$y_{ab1}, y_{ab2}, \ldots, y_{abn}$	$y_{a..}$	$\bar{y}_{a..}$
Totals		$y_{.1.}$	$y_{.2.}$	…	$y_{.b.}$	$y_{...}$	
Averages		$\bar{y}_{.1.}$	$\bar{y}_{.2.}$		$\bar{y}_{.b.}$		$\bar{y}_{...}$

So, this is the data table, which tells factor A along this direction, factor A at setting 1, setting 2 so on to the ath setting of factor A. Similarly, for factor B you move horizontally, you have first level of factor B, second level of factor B so on to b levels of factor B. If it is strictly 2 factorial design, you will stop for A at 1 and 2, for B at 1 and 2, so you will be having only this combination, these 4 elements.

Obviously, within each cell 1, 1, you can have n number of repeats, so i j k; i is 1, j is 1, k is 1, i is 1, j is 1, k is = 2, i is 1, j is 1, so 1 to k = n, so you have n entities in this first cell. In the second cell, you are having the same level of factor A at 1 but now you have gone to the second level of factor B, again you can have n repeats, you are all assuming that the repeats per cell is constant, there is the number of repeats done here is equal to the number of repeats done here and anywhere else in any other cell.

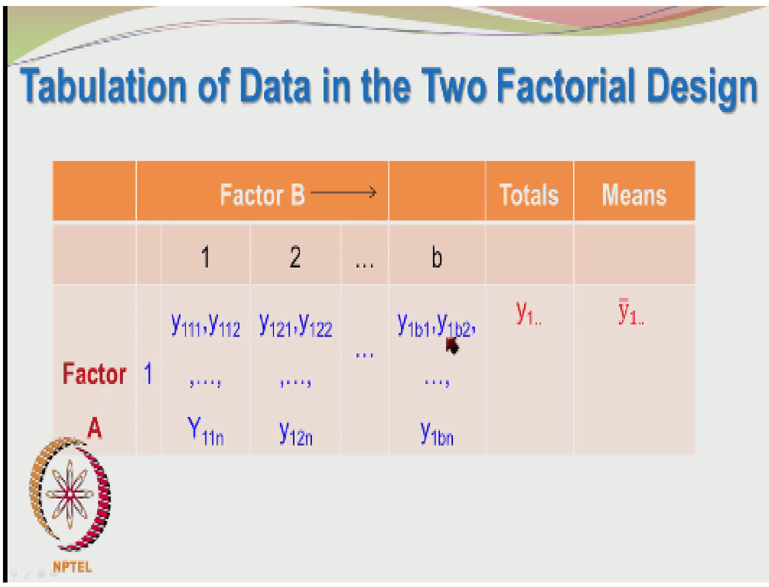
(Refer Slide Time: 33:39)



So, you have factor A at level 1, factor B at level 2, first repeat, second repeat so on to nth repeat, the same logic is applied to all other cells in this table, so I have made it a bit simple, you can see that only factor A is changing from first level to 8th level, factor B is at kept at first level, j is equal to 1, second index is equal to 1, so you have y1 corresponding to level 1 here, B is at level 1, so B is always going to be 1 and first repeat.

And then, it is 1, 1 second repeat, so 1 to 1, 1, nth repeat. Similarly, when you go to the next level of factor A, the A index i will take a value of 2 and so you will have 2, B at still at level 1; 211, 212 so on to 21n, so when you add up all these things, you get y dot 1 dot, which means only factor B is kept at level 1 and the others are all added up and then you can take the average y dot 1 dot that means j is unaltered, j is kept at one setting but both a and n are varied.

(Refer Slide Time: 34:55)



So, it will be a * n, we will come to that very shortly but I hope you have understood the layout of this particular table. This is another form, where you are now fixing factory A at one level but looking at the different levels of factor B, the notation goes as y1, j is =1, first repeat and this one would be y, first setting of factor A, B is kept at the level 2 here throughout and the first repeat, again 1 stands for factor A setting i.

(Refer Slide Time: 35:39)

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

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

Index i is = 1 and then factor 2 is kept at level of 2 and then first repeat, so again, 122, so 1 to y1 2n. Similarly, you can do for the bth level of factor B. So, now this is very important, it tells you how to do the summing and averaging, y dot dot means i is kept constant and j and k are summed upon and j and k are summed upon for b levels of j and n levels of k, you have Yijk. When you want to take average, you simply divide the total sum given here by b * n.

For, this is applicable for every value of i, 1, 2, so 1 to a, y dot j dot; now you are keeping the index j as constant and then you are adding over ith index and also the kth index that is what is represented here over a, n elements and so you have y bar dot j; dot is = y dot j dot / a * n, where y dot, j dot is nothing but this summation and j can go from 1, 2, so 1 to b, please spend a bit of time trying to understand all these things, it will help you later and it will not cause any confusion subsequently.

(Refer Slide Time: 36:48)

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij.} = \frac{y_{ij.}}{n} \quad k = 1,2,\dots,n$$

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

Now, when I do Yij dot, I am saying that only; I am adding over k, I am keeping i and j at constant values, so only case varied from 1 to n, you have Yijk summation to give Yij dot, y bar ij dot means whatever sum I have got here, I am going to divide by the number of elements in that particular sum, which is nothing but n and I am getting that y bar dot j dot average value, Y dot dot dot means, I am summing over all the index; indices rather i, j and k for the response Yijk and so I get this particular expression.

(Refer Slide Time: 37:50)

Formulae for Sum of Squares

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{...}\right)^2 = bn\sum_{i=1}^{a}\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2 + an\sum_{j=1}^{b}\left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2$$

$$+\, n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{ij.}\right)^2$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

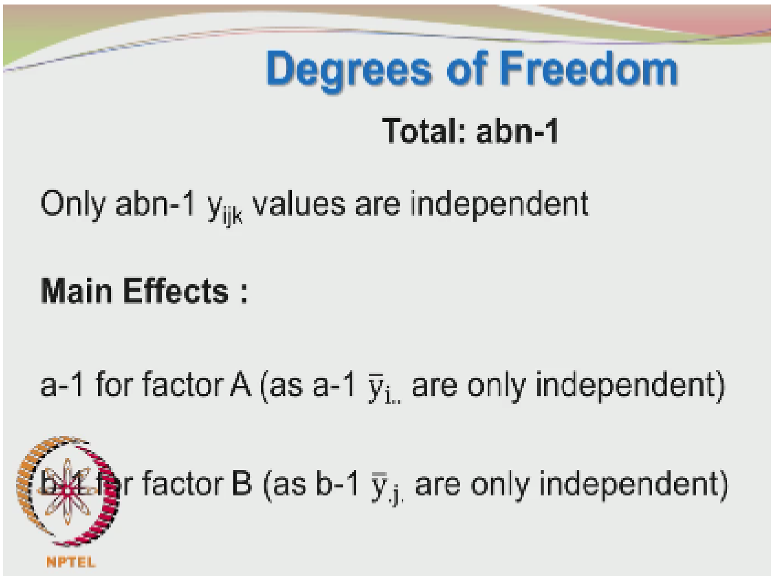
NPTEL

Now, I divide it by a, b, n; a, b, n is the total number of terms, I get y bar triple dot, I hope you understood the nomenclature, this is very important. If you do not understand, please try to understand using the current lecture material and discussions because otherwise you are going to get hopelessly confused, right. Now, this is very fascinating, you can see that the solution of the total sum of squares.

I told you the total sum of squares is Yijk, individual observation response - y bar dot dot dot okay and so you have Yijk - y bar dot dot dot individual observation - a grand average that is squared and then that is done for all the elements in that table, which we saw, we are doing it for all the elements, we take the grand average and then we subtract the grand average from each and every entity given in this table.

And then, that is sum of squares will be a total sum of squares and that is resolved into the sum of squares due to factor A, which is the ith treatment mean corresponding to factor A from the overall grand average. Similarly, this is the treatment mean corresponding to the factor B and that you subtract with the overall grand mean and this is the interaction effect between A and B and this one would be the individual observation within each cell subtracted by the average of those repeated observations.

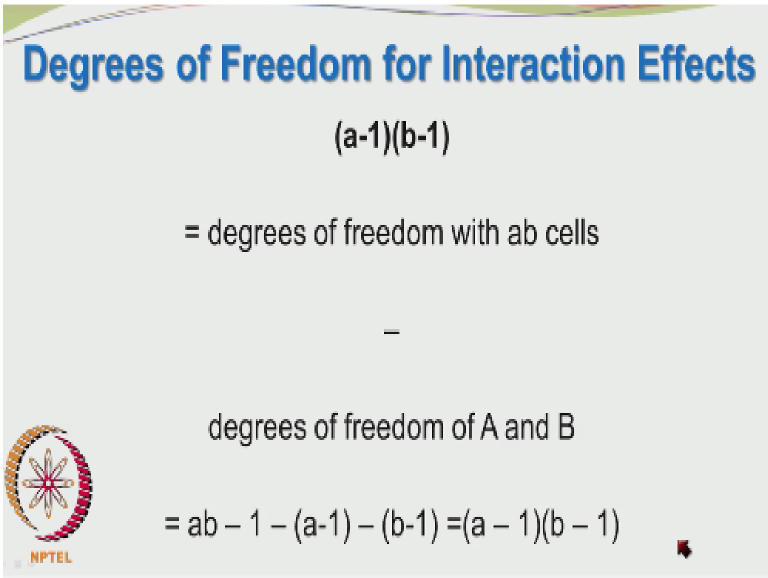
(Refer Slide Time: 39:46)



Again, this is very similar to the single factor experimentation, where explain these in more detail, same concepts apply also here. So, this is the effect of A; sum of squares of factor A, sum of squares of factor B, sum of squares of interaction AB, this is slightly difficult to understand and the sum of squares of the error. So, when you look at the main effects, you have only a - 1 independent entities.

Because out of the y bar i dot dot treatment means, there are A such treatment means and only a - 1 of them are important. How do you get y bar i dot dot that is quite simple, so each of these entities here represents y bar i dot dot, here i is = 1, here i is = 2 and here i is = a, so you have y bar 1 dot dot, okay and you have y bar 2 dot dot so, 1 to y bar a dot dot but since you are using all of these, if I take average of all these divided by a, I will get the grand mean.

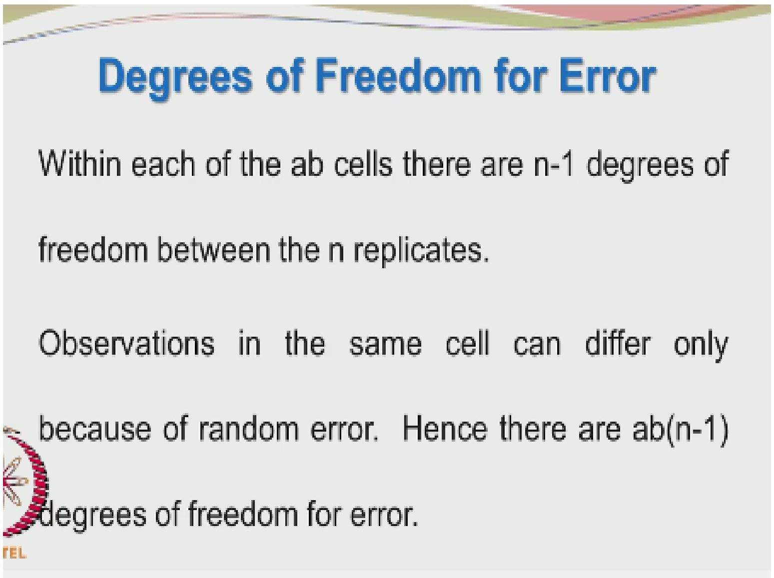
(Refer Slide Time: 41:33)



So since, I am taking grand mean from these only a - 1 are independent, same argument can be given for factor B, these are the treatment averages for factor B and I can get the grand mean from averaging out these treatment B means, so only B - 1 of them are important okay, so that is very important. So, we have finished to the main effects, what about interaction A and B. If interaction A and B slightly difficult to understand.

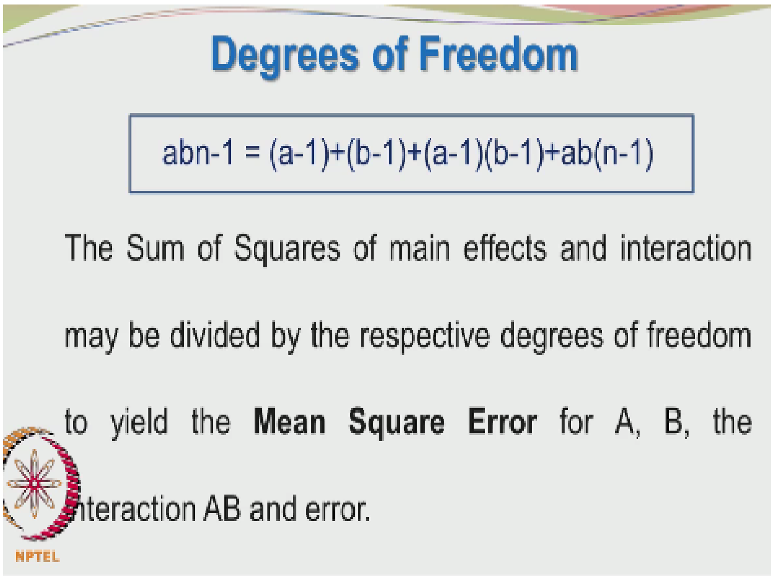
Factor A is having a degree of freedom of a – 1, factor B is having a degree of freedom of b – 1, so the interaction of ab would be a – 1 * b -1. If you look a bit further Montgomery in his book on design and analysis of experiments has given a nice explanation. What he says is first see the degrees of freedom with ab cells, the degrees of freedom of the ab cells would be ab - 1 and then you subtract from this available degrees of freedom.

(Refer Slide Time: 42:29)



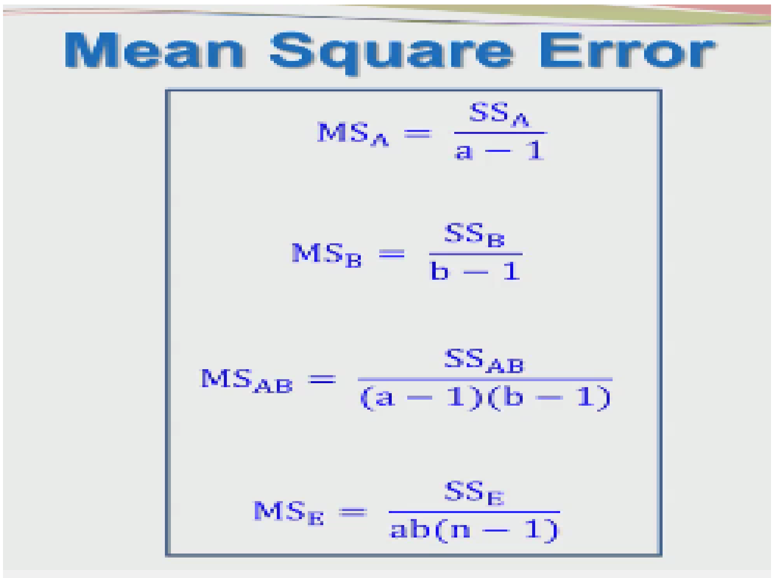
The degrees of freedom due to factor A and the degrees of freedom due to factor B and you will get nicely ab – 1 - a - 1 – b -1, which is nothing a – 1 * b – 1. So, for the degrees of freedom for the error, you are having ab cells totally but within each of these ab cells, even though you are having n repeats only n - 1 of them are important. So, totally you have ab * n - 1 independent entities and that is the degrees of freedom for the error.

(Refer Slide Time: 43:22)



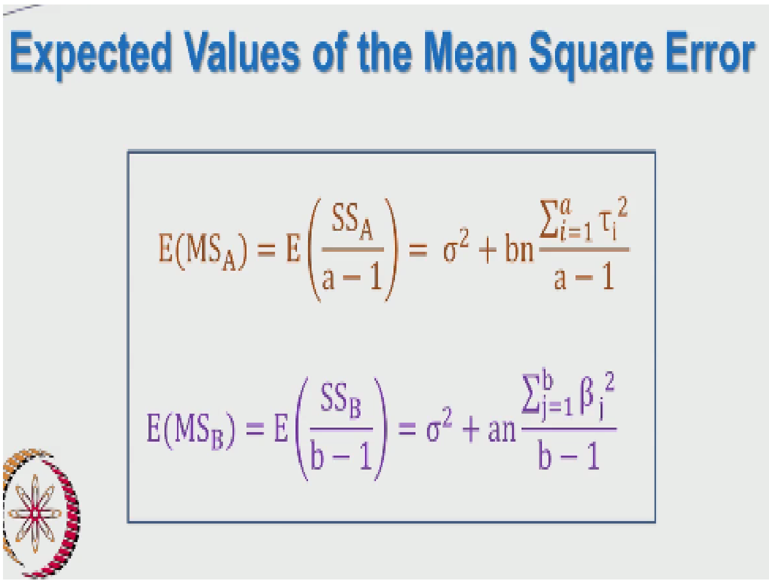
Again the argument here is very similar to what we did for single variable experimentation, if you are finding this slightly difficult to understand; I request you to please go back to the single variable experimentation design discussion and now whatever we are doing will be easy to follow. So, the degrees of freedom may be partitioned into ab n – 1, which is a total degrees of freedom that may be partitioned off into a - 1 for A, b - 1 for B, a -1 * b - 1 for AB and ab * n - 1 for the error.

(Refer Slide Time: 43:59)



Now, we have seen the total sum of squares for each case, these are for a, for b, ab and for error, we can divide each of these sum of squares with the degrees of freedom and we can get the mean square error, we can get the mean square not the mean square error, sorry. So, mean square A would be sum of squares of A/ a – 1, mean square B would be sum of square of B/ b – 1, mean square AB is sum of squares of AB/ a – 1 * b – 1, mean square error would be sum of squares of E by the degrees of freedom for error, which is ab * n – 1.

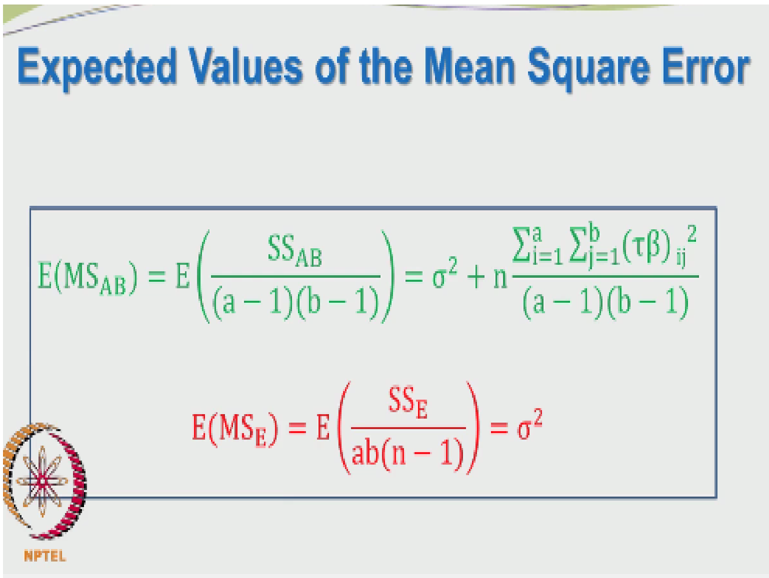
(Refer Slide Time: 44:22)



Again, this is very similar to what we had done with a single factor experimentation, expected value of the mean square of A is the error variance Sigma squared + the additional variability cost by the 8th factor; the 8th or the first factor a is contributing to the response, it cannot be ignored and so at least, one setting of the 8th factor is having an effect and so you are adding up all the treatment effects here, so this is over and above Sigma squared.

If this were not important, then you will have only Sigma squared, if this term vanishes or becomes negligible, then this will have only Sigma squared, so the variation due to changing factory A may then be attributed to random noise. Similarly, expected value for mean Square B would be Sigma squared plus a * n Sigma j equals 1 to b beta j squared/b – 1, when these effects are insignificant, then the variation due to B is because of random error but if it is not so, factor B is effective in altering the response of the process, then you have this additional contribution.

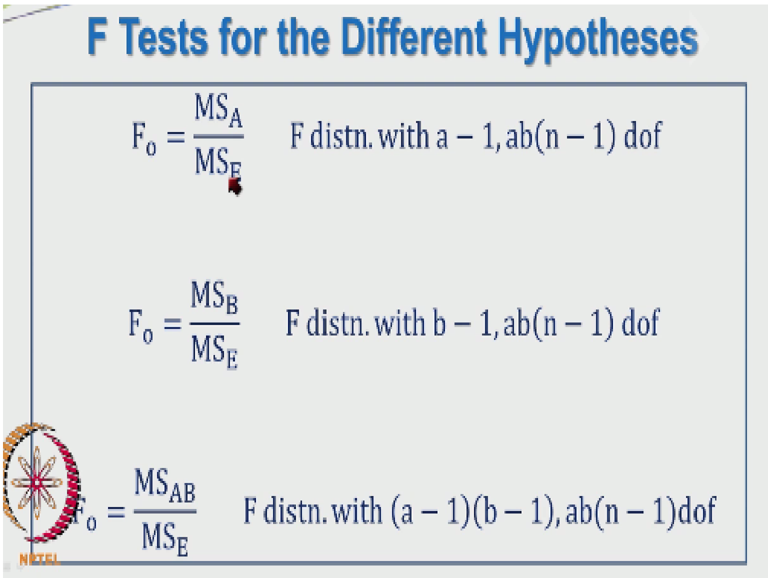
(Refer Slide Time: 45:38)



Similarly, for AB, you can see the nice symmetry in all these things, you can see now for AB, I am doing the summation over both, i equals 1 to a, j equals 1 to b, tau beta ij whole squared, expected value of mean square error would be expected value of sum of squares of error by ab * n – 1. So, here there is no inclusion of any other term, we are talking about pure error only. I hope you have understood this discussion; this is very straight forward.

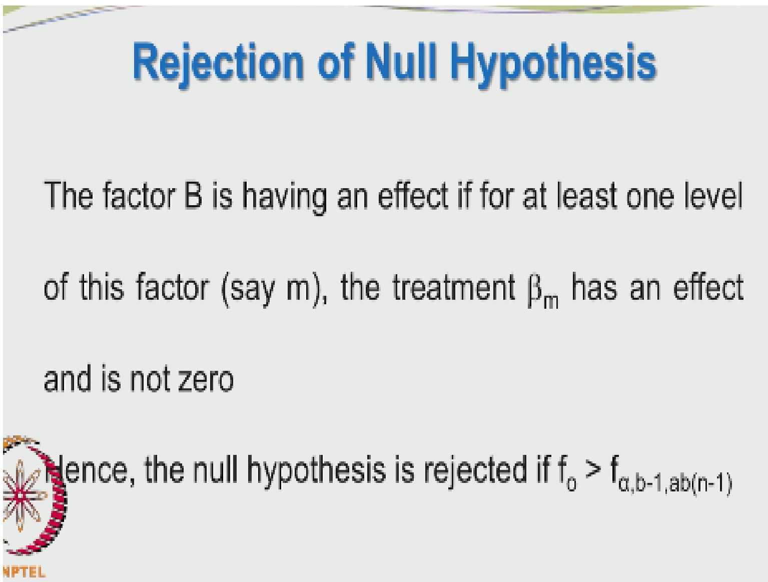
But this expected value is only to tell that you are having the error variance and additional contribution variance or variability due to the 2 main factors and their interaction. If the main factors and or the interactions were ineffective, then that variability are also estimates of the random error component. If the effects are significant, then you cannot take it as random variation but you have to separately account for the effects of those factors.

(Refer Slide Time: 46:47)



But this is pure error and so you are getting Sigma squared. Now, we can do the F test, what we do is; we do the mean square for A. How do we find means square for A? Sum of squares of A/ a- 1, we already saw how to find sum of squares of A, so means square A/ mean square error. Similarly, we do the mean square B/ mean square error means square AB/ mean square error. We have seen the beauty of the F distribution where it has numerator degrees of freedom and denominator degrees of freedom.

(Refer Slide Time: 47:50)

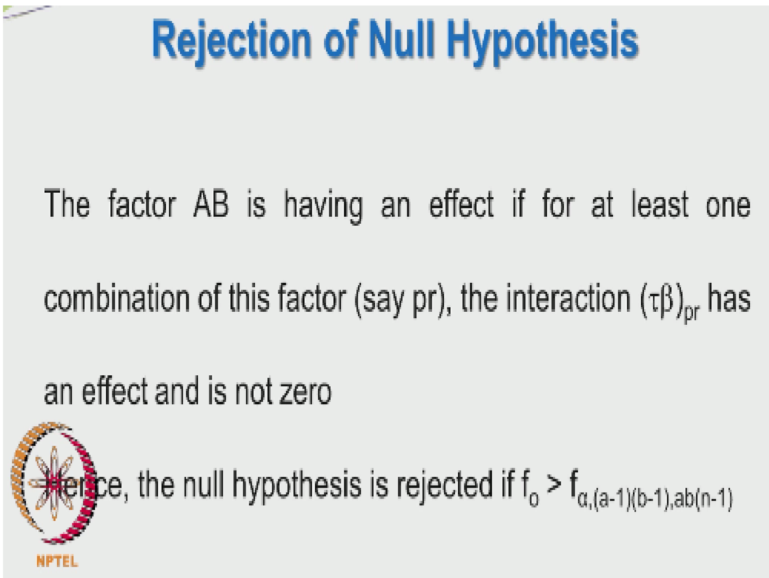


Numerator degrees of freedom for A would be simply a -1, denominator degrees of freedom would be ab * n – 1. Similarly, for factor B, it is b - 1 degrees of freedom in the numerator and ab * n – 1 degrees of freedom in the denominator. Similarly, for interaction AB, you are having a – 1 * b - 1 degrees of freedom for means square AB / ab * n - 1 degrees of freedom for the denominator. So, with this in mind, we have to either accept or reject the null hypothesis depending upon the value of F.

We see the value of this F and see whether it falls in the rejection region or in the acceptance region. It falls in the rejection region, if the statistic F0 comes to F alpha, which is the level of significance; alpha is the level of significance 0.5 is usual, you may also use 0.25 or 0.01 or even 0.1, a - 1 is the numerator degrees of freedom, it is the degrees of freedom for factor A, ab * n - 1 is the degrees of freedom for error.

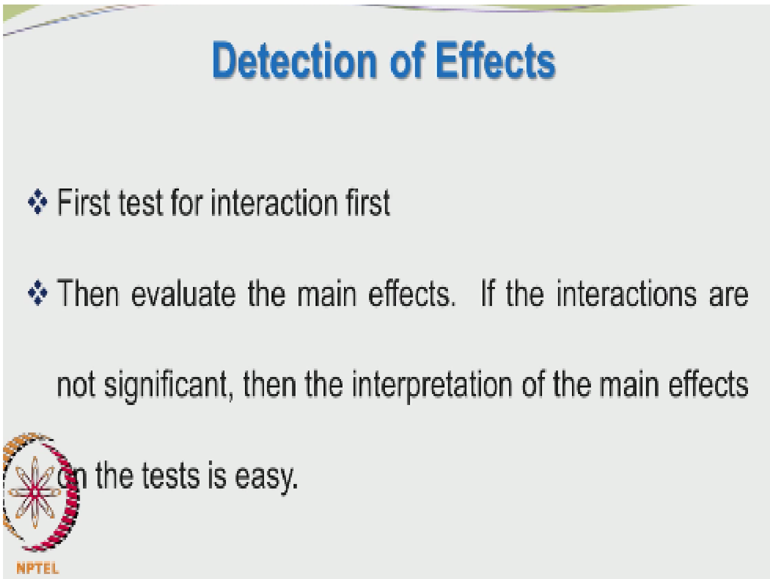
So, if you find this critical value and your statistic is much higher than that and it falls in the critical region, you reject the null hypothesis. Similarly, you do the same thing for factor B, find out F alpha b - 1 numerator degrees of freedom ab * n - 1 denominator degrees of freedom for the error component, if F0 is > F alpha b -1 ab * n – 1, so be it, so you are able to reject the null hypothesis that even factor B is not significant.

(Refer Slide Time: 49:13)



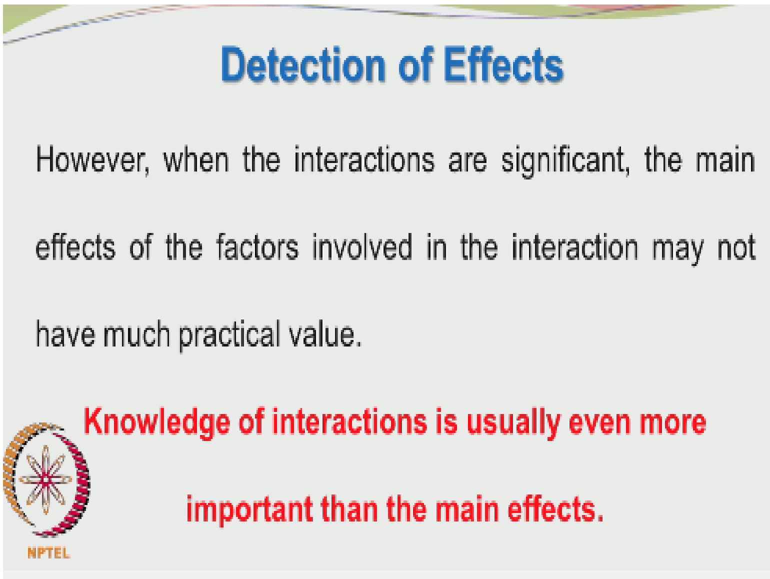
You have to state that even factor B is significant, same thing you do for AB. Here, you use alpha level of significance a – 1 * b - 1 degrees of freedom for the interaction AB and that is the numerator degrees of freedom ab * n - 1 is again the denominator degrees of freedom corresponding to random error and if this F value is exceeding the critical; if this F value is exceeding the critical F value given according to this relation, then you reject the null hypothesis that the ab interaction is insignificant.

(Refer Slide Time: 49:49)



So, you can test for the interaction effects first and then look for the main effects. If the interaction effects are not significant, interpretation of the main effects on the test is quite simple.

(Refer Slide Time: 49:59)



Again, as I said previously before interaction effects are more important than the main effects.

(Refer Slide Time: 50:08)

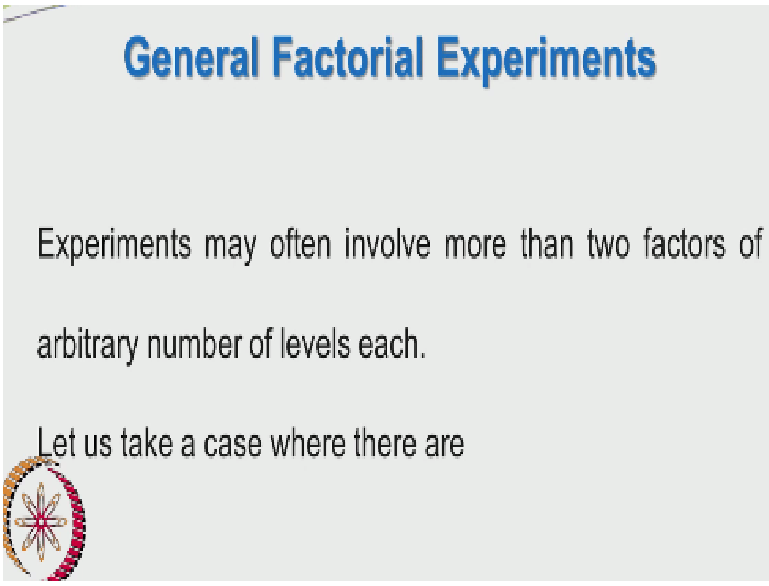
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F_0
A Treatments	SS_A	a-1	$SS_A/(a-1)$	MS_A/MS_E
B Treatments	SS_B	b-1	$SS_B/(b-1)$	MS_B/MS_E
Interaction	SS_{AB}	(a-1)(b-1)	$SS_{AB}/(a-1)(b-1)$	MS_{AB}/MS_E
Error	SS_E	ab(n-1)	$SS_E/ab(n-1)$	
Total	SS_T	abn-1		

So, this is the summary of the ANOVA table, sum of squares of A, sum of squares of B, sum of squares of AB, sum of squares of error by sum of squares of total sum of squares. If I add up all these elements, I will get total sum of squares. Here the degrees of freedom was a -1, b – 1, a – 1 * b – 1, ab * n – 1, we have already seen this several times in the past, so I am not going to spend too much time on that and when you add up all these degrees of freedom, you will get ab 8 n – 1.

Mean squares are formed by dividing the respective sum of squares or dividing the sum of squares by the respective degrees of freedom, you will get SSA/ a – 1, SSB/ b – 1, SSAB/ a – 1 * b – 1, SSE/ab * n -1. What are the terms in the denominator in each of these expressions? They are nothing but the degrees of freedom associated with that individual factor or the combination of factors.

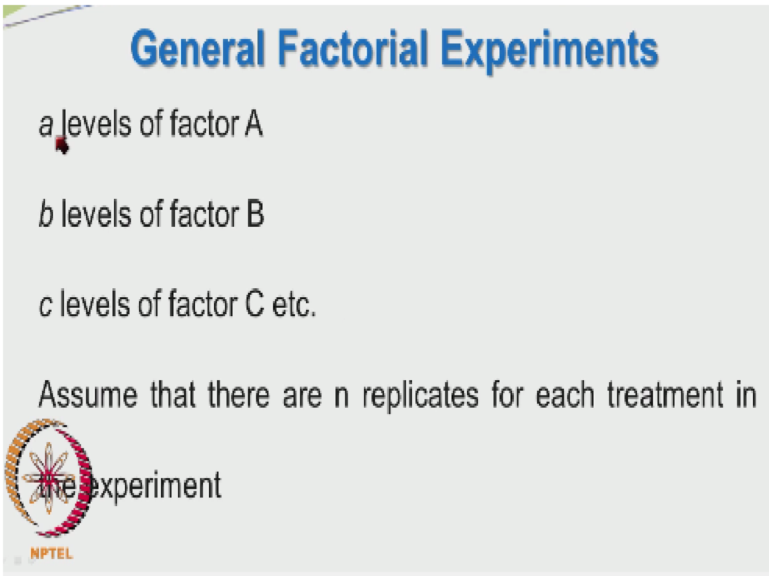
F0 is defined simply as sum of squares of A/ a – 1, sum of squares of B/ b – 1, so all these terms are individually divided by mean square, even interactions, you do not leave it alone, you take the sum of squares of AB divided by the degrees of freedom for AB, you get the mean squares for AB that you divide by the mean square error, so you can see which of these are falling in the critical or rather the which of these F values are exceeding the critical value.

(Refer Slide Time: 51:54)



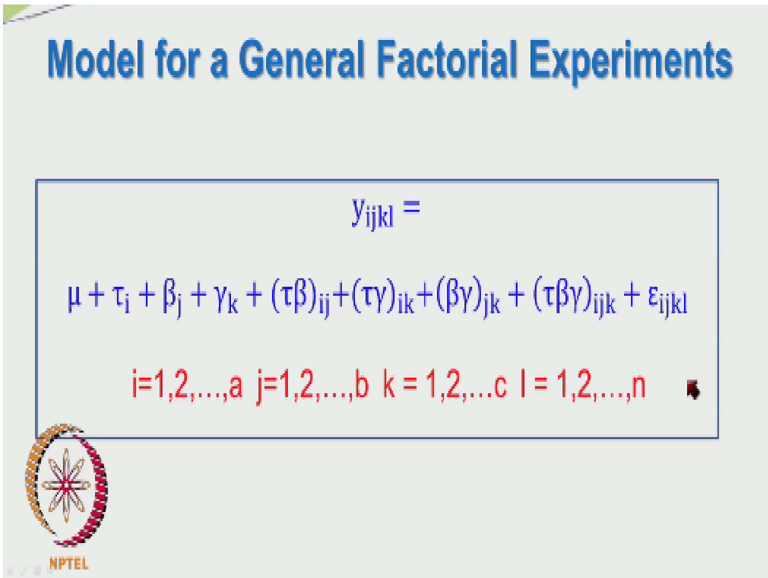
And hence falling in the rejection region based on which you can state your acceptance or rejection of the null hypothesis. Now, we move on to general factorial experiments, the concept is very similar to the 2 factorial experiments, you can conduct with any number of factors with any number of arbitrary levels, factor A can be 2 levels, factor B can be 2 levels, factor C can be 4 levels and each may have a certain number of repeats.

(Refer Slide Time: 52:21)



We assume that the number of repeats is constant. So, I will go through it very quickly, so you can have in general, a levels of factory A, b levels of factor B, c levels for factor C and we have n equal number of replicates or repeats, we have equal number of replicates or repeats for each treatment in the experiment.

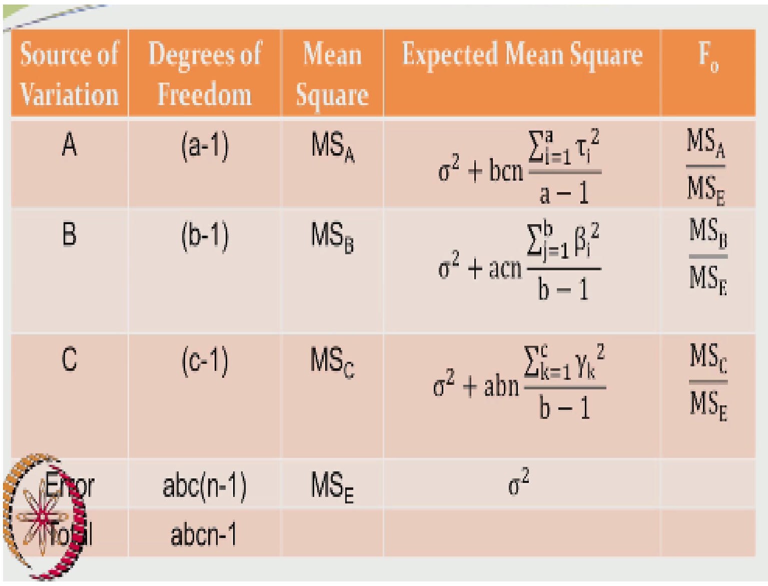
(Refer Slide Time: 52:39)



So, we have this model Yijk, so you have an additional subscript now, i stands for factor A, the the index i stands for Factor A, index j stands for factor B, index k stands for factor C and l stands for the repetition. This is the overall mean, tau i represents the effect of the ith level of factor A, beta j means the effect of the jth level or setting of factor B, gamma k refers to the effect of the kth setting or level of factor C.

And this represents the interactions; binary interactions between factors A and B, binary interaction between factory A and C, binary interaction between factor B and C.

(Refer Slide Time: 53:57)



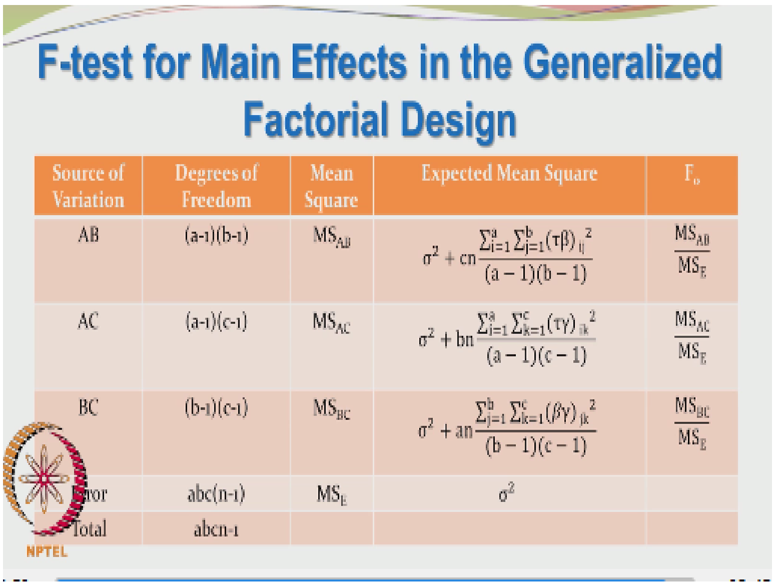
Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square	F_0
A	(a-1)	MS_A	$\sigma^2 + bcn\frac{\sum_{i=1}^{a}\tau_i^2}{a-1}$	$\frac{MS_A}{MS_E}$
B	(b-1)	MS_B	$\sigma^2 + acn\frac{\sum_{j=1}^{b}\beta_j^2}{b-1}$	$\frac{MS_B}{MS_E}$
C	(c-1)	MS_C	$\sigma^2 + abn\frac{\sum_{k=1}^{c}\gamma_k^2}{b-1}$	$\frac{MS_C}{MS_E}$
Error	abc(n-1)	MS_E	σ^2	
Total	abcn-1			

And this is the ternary interaction between factors A, B and C, then you have the error component, you can see that the index i varies from 1 to a, index j varies from 1 to b, index k varies from 1 to c and the index l varies from 1 to n, right. So, we have again this table, you have the source of variation, the degrees of freedom are a -1, b – 1, c – 1 and abc * n – 1, you may be asking what happens to the interactions, AB, BC, AC.

Please be a bit patient it is soon coming and the mean square error is the sum of squares contribution of A/ the degrees of freedom for A, which is a – 1, calculating the sum of squares from now on will become more tedious because there are many factors, it is not expected that you do these things either with the hand calculation or even with spread sheet. There are software's like Minitab, which is available and you may resort to such software to calculate the treatment squares.

And the expected mean square value would be nothing but the random error component plus the contribution by the individual treatments. So, you have Sigma squared by the plus the contribution from the treatment a, Sigma squared + contribution from treatment B Sigma squared + contribution from factor C and so on. If these contributions are not there, then the expected mean square would be an estimate for the random error component sigma squared.

(Refer Slide Time: 55:30)



F-test for Main Effects in the Generalized Factorial Design

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square	F_0
AB	(a-1)(b-1)	MS_{AB}	$\sigma^2 + cn\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
AC	(a-1)(c-1)	MS_{AC}	$\sigma^2 + bn\frac{\sum_{i=1}^{a}\sum_{k=1}^{c}(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$\frac{MS_{AC}}{MS_E}$
BC	(b-1)(c-1)	MS_{BC}	$\sigma^2 + an\frac{\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$\frac{MS_{BC}}{MS_E}$
Error	abc(n-1)	MS_E	σ^2	
Total	abcn-1			

Otherwise, you have not only sigma squared but an additional contribution from the main effects or as we will see the interaction effects, so the mean square error is sigma squared. Now, you have the interactions and again you have the sigma squared error variance plus the contribution from AB interaction, AC interaction and BC interaction and you can always form the means square and you can divide the mean square for the effect or the interaction with the mean square error to get the F0 value.

(Refer Slide Time: 56:18)

F-test for Main Effects in the Generalized Factorial Design

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square	F_0
ABC	(a-1)(b-1)(c-1)	MS_{ABC}	$\sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$\frac{MS_{ABC}}{MS_E}$
Error	abc(n-1)	MS_E	σ^2	
Total	abcn-1			

Then, you can identify or create the critical value based on the numerator and denominator degrees of freedom and then you can see whether the F value is lying in the acceptance region or in the critical region, then suitably you can accept or reject the null hypothesis. So, for example, ABC, you are having a -1 * b -1 * c - 1 as the degrees of freedom and the mean square is the sum of squares of ABC/ a -1 * b -1 * c - 1 and the expected mean square here, we do not use it in the F test.

It is just to show that it has an error variance sigma squared plus the contribution variability because of the ABC interaction and to find the F0 value, you divide mean square ABC by mean square error, you have numerator degrees of freedom as a – 1, b -1, c – 1, denominator degrees of freedom has ABC * n – 1, you can look up the F probability charts based on these degrees of freedom and see whether the F statistic is lying in the critical region.

So, this completes our discussion on factorial design, the 2 power K factorial design is quite useful, elegant and easy to understand for small number of factors. What is to be done when there are large numbers of factors, even then the number of experiments may blow up including the repeats, it may be quite an investment to do so many experiments, there are some elegant alternatives to the general factorial design.

So, what I am trying to say is the factorial design is not only restricted to 2 levels, it can be generalized into any number of factors and any number of levels, the only problem is even when you generalize, you also have to keep tab on the number of experiments you have to do and coupled with the repeats, general factorial design involving many levels and many factors and repeats will lead to again a large number of experiments.

So, we have to find alternatives to even these factorial designs and this will form the background for our future discussions. Thank you for your attention.