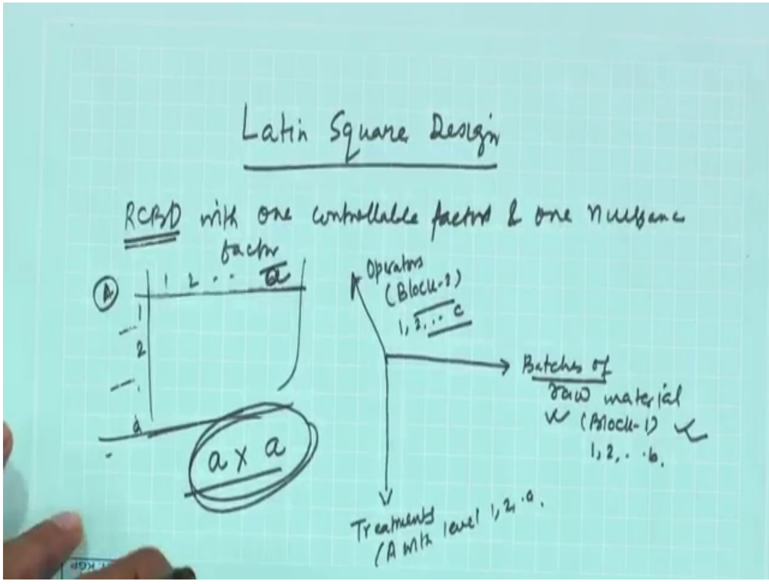
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 28
Randomized Complete Block Design (RCBD): Latin Square Design

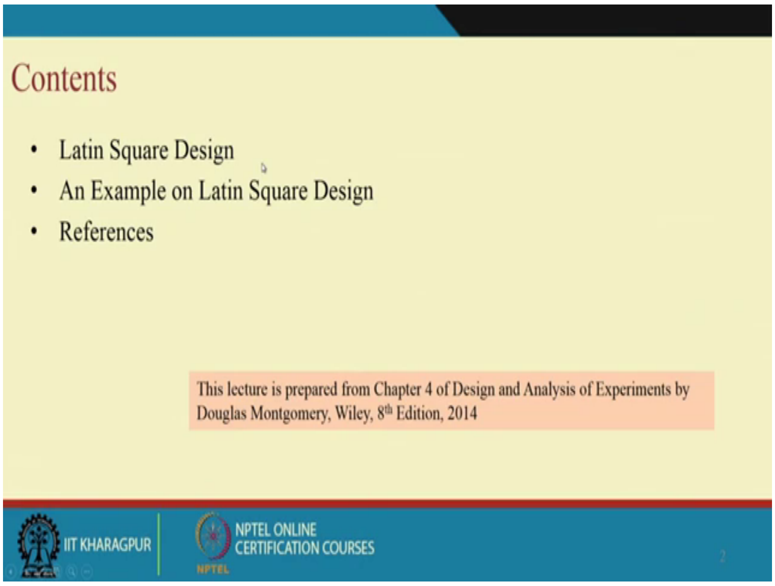
Welcome. We continue Randomized Complete Block Design. In this lecture, we will discuss Latin square design.

(Refer Slide Time: 00:23)



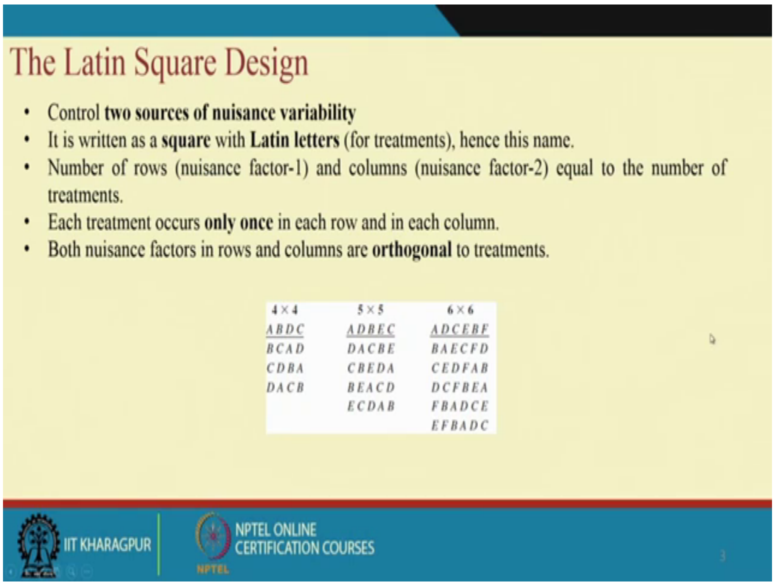
Let us see the contents. We will first explain what Latin square design is.

(Refer Slide Time: 00:34)



Then we show you the statistical analysis part of Latin square design, then one example, in fact, 2 examples we will see and references. The text book from where we have taken the material is Design and Analysis of Experiments by Montgomery.

(Refer Slide Time: 01:03)



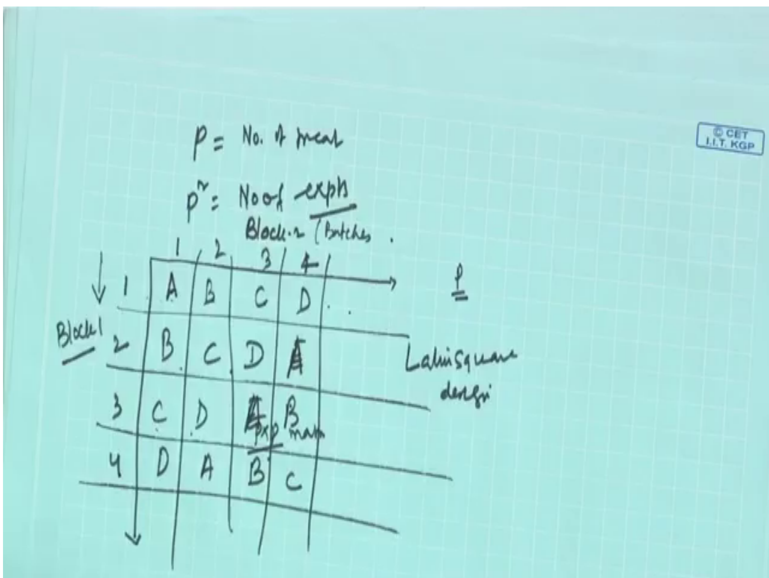
What is Latin square design? In randomized complete block design, with one controllable factor and one nuisance factor, you have seen that you have this kind of things 1, 2, a and 1, 2, b here so many blocks and then different levels and then we have seen the statistical analysis all those things. So, that means, in the first lecture of RCBD, I have shown you that there is one controllable factors A and there is blocking or nuisance factor 1 and which is having the b level, for example; batches of raw materials in that case b number of batches. Now, suppose instead of these 2 that mean that batches of raw material this is your block 1, and this side suppose you have the treatments, the factor A, with level a, level 1, 2, a. Suppose, there is another block maybe the operators, it is block 2.

So, these batches of raw material let it be 1, 2, b and operators also you can think of 1, 2, c. You can go for statistical model that y ijk and then what will happen? Treatment effect, batch 1, block 1 effect, block 2 effect then errors all will be there. But, if you do this what will happen you require more number of experimental runs and blocking in this manner also is costly.

Instead of this, suppose, you create a special kind of design in such a manner that, if you have A number of treatments you chose a number of blocks for blocking variables block 1 and block 2. Then, it will give you a square matrix kind of thing. Suppose, if a is the number of treatment, and then if I go for a number of blocks for a particular blocking variable then you will be getting this side a and this side a, 1 a by a square matrix, this takes care of the first block.

Then what will happen to the second block and how do you add up this one without increasing the number of experiment. This is a special kind of things.

(Refer Slide Time: 05:30)



Now, I will show you these that how the other block is also accommodated here, and then the treatments are combined within the blocks to group of block to blocking the effectors in such a manner that your A cross A or P cross P, suppose P is the number of blocks for the number of treatments, then you will with P square number of experimental runs you will be able to estimate the treatment effects.

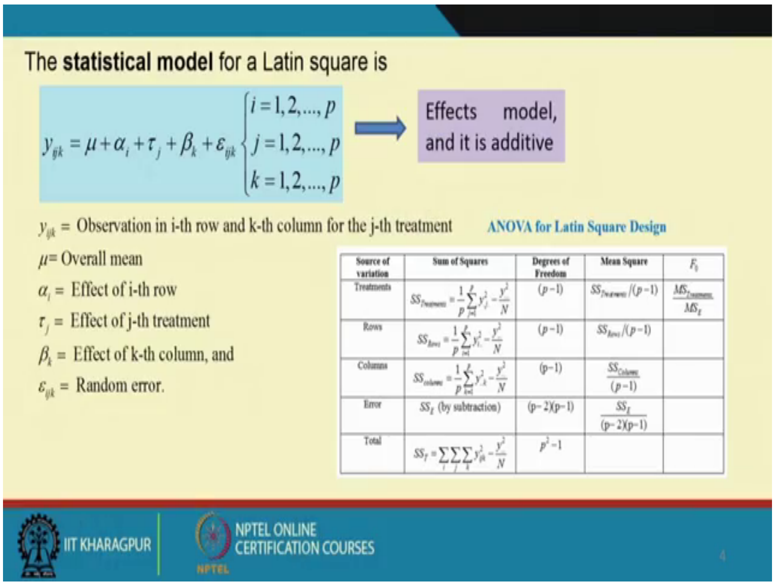
This kind of design is coming here. You see the first one it is a 4 cross 4, ABCD another treatment and then suppose, the row is block 1 and column is block 2 and in between the cells are basically assigning the treatment for the 2 different block combinations. If there are 5 numbers of treatments, you take 5 blocks for first blocking variable factors and you can run the 5 blocks for second blocking factors, 5 levels or if there are 6 treatments, 6 levels each of the 2 blocks.

In that case, you will create a square matrix where rows represents the level of block 1, column represents the level of block 2 and in between the Latin words represent the treatments. That is what is written here in this kind of design is known as Latin square design. Latin square design control 2 sources of nuisance variability it is written as a square with Latin letters. ABCD although these are Latin letters and each letter represent a particular treatment level.

Number of rows, there will be again number of rows equal to the number of treatments and that represents that nuisance factor 1, number of columns equal to the number of treatment represent the nuisance factor 2 and what happen in each treatment occurs only once in each row and each column combinations. Each row and each column, only one time that treatment will occur. For example; if I consider 5 by 5 the second one, 5 by 5 this Latin square you see that if you consider first row A occurs once there is no more A, similarly if you consider first column A appears once. If you consider suppose the fifth column D appear once, as well as fifth row also D appear once. Each treatment occurs only once in each row and in each column.

Another important thing is that this nuisance factor in rows and column they are orthogonal to the treatments. So, that is another important consideration.

(Refer Slide Time: 08:43)

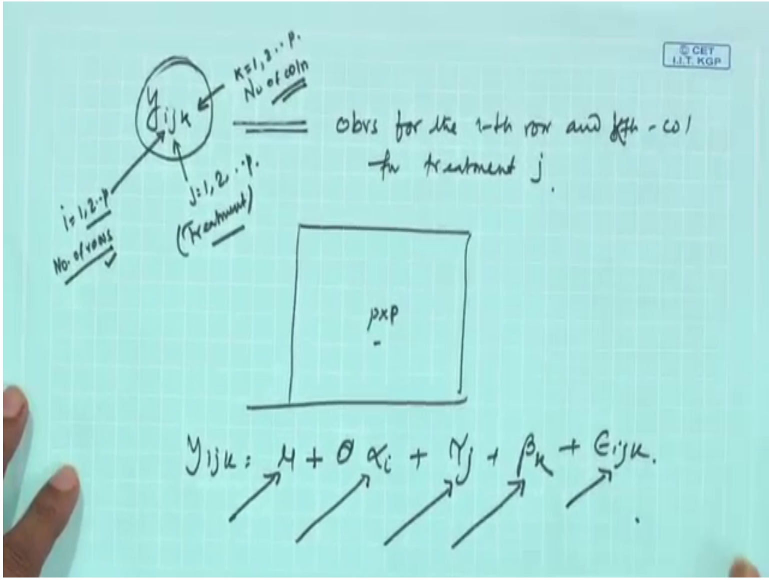


This type of design is known as Latin square design. What is Latin square design? Latin square is a squared ultimately it control nuisance factor, 2 nuisance factor, and there if there are 5 numbers P number of treatments. So this will be a P cross P matrix. Each of the cells represents A treatment.

Suppose ABCD, then BCDE, then CDE. So, ABCD, BCDA then CDAB then DABC through its batches of raw material, then raw material batch 1, batch 2, batch 3, batch 4, if it is operator 1, operator 2, operator 3, operator 4. That mean the number of the size of the Latin square will be determined by the number of treatments, then there are 4 treatments. There are 4 cross 4 matrix, and each of the rows every treatment will occur once. Similarly, in each of the column every treatment will occur once; this is our Latin square design.

So, what happen? You have 3 different factors; one is treatment and 2 nuisance factors, but, you are in a position that you just conduct P square number of experiments and then you are able to block the 2 nuisance factor, as well as you are able to estimate the effect of the treatments and also you are in a position to conclude whether the treatment levels are differently effecting the response variable or not.

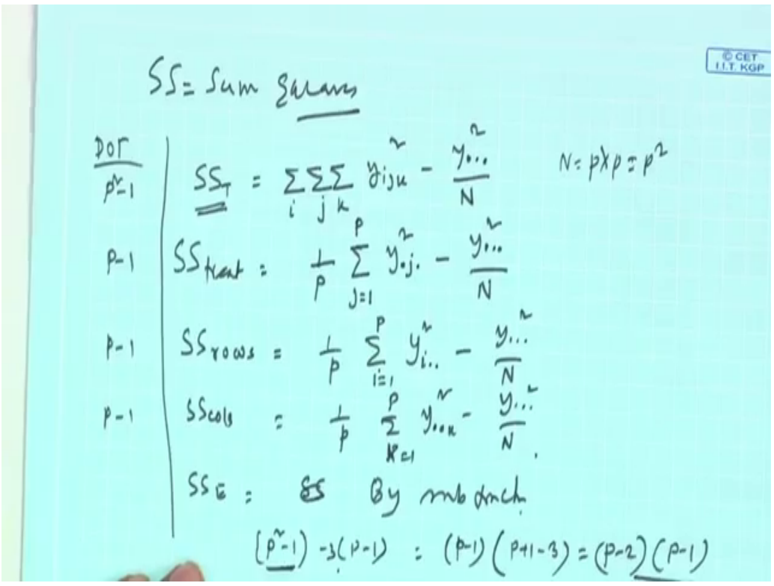
(Refer Slide Time: 11:14)



Now what is the statistical model? The statistical model it has 3 directions y ijk. i stands for i equal to 1,2, P, that is, number of rows. k stands for k equal to 1, 2, P this is number of columns. That means i for blocking factor 1, nuisance factor 1 and this k for nuisance factor 2 and what is j? j also 1, 2, P, but this is for treatment. Then what is y ijk observation on i-th row and k-th column for the j-th treatment. This represent observation for the i-th row and j-th column for treatment k-th column for treatment j. That is what your general observation y ijk is. You have y i 3 directions, but in the data set you have only P cross P. I will and P cross P data is able to do our purpose this number of data able to do our purpose.

What will be our model then? y ijk this will be mu plus alpha I plus tau j plus beta k plus epsilon ijk. Overall mean block 1 effect, treatment effect, block 2 effect, error and sources of variability that is why 2 blocks, 1 treatment, 1 error being orthogonal there is no interaction effects.

(Refer Slide Time: 13:54)



Now, how to compute the SS sum square? This computation is very important. Let us see the table. This table, right hand, side lower most this one; this is ANOVA table for Latin square design. Our sources of variations are treatment, rows, column, error these 4 source; treatment, row factor, column factor, error.

Now, SS treatment you will be computing like this 1 by p j equal to 1 to p and if j varies y dot j dot square minus this and it is degree of freedom is p minus 1. SS also, row will vary and SS column that column factor will vary and total all this things you will consider. I will write down again. SS total you have 3 direction, one is your k then ijk, y ijk square minus y dot dot dot, this is grand total by N. What is N here? N equal p into p square. N equal to p cross p equal to p square. Suppose, you want treatment, what will happen? There are p number of cases, so, your j will vary from 1 to p, then your this side y dot j dot this square minus y dot square by N, j is changing.

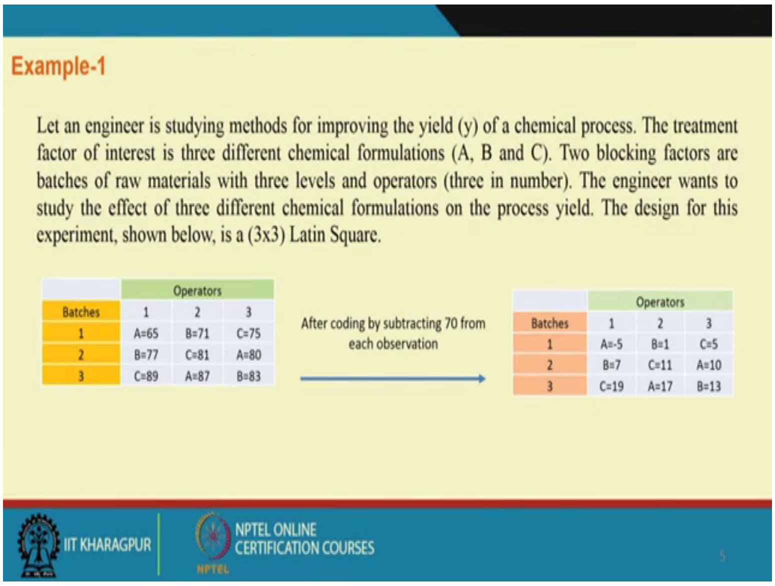
SS row if I write rows this is row blocking, then this will be i equal to 1 to p, y square i dot dot minus y dot dot dot square by N. Similarly, SS columns if I write 1 by p, column variable will change that is k equal to 1 to p, y dot dot k square minus y dot dot dot square by N and SS error by subtraction. What will be the degree of freedom for each of these sources? Here, p square minus 1 is the degree of freedom for the total for treatment p minus 1, for rows p minus 1, for column p minus 1, then what will be for SS E? p square minus 1, minus p minus

1, p minus 1, p minus 1, 3 p minus 1. If you take common p minus 1 this will be what, p plus 1 into p minus 1, so, p plus 1 minus 3, p minus 2 into p minus 1.

As a result, you see that from here the degree of freedom is p minus 1 p minus 1 p minus 1 p minus 2 go to slide p minus 2 and p minus 1 and p square minus 1, and calculate mean square calculate F, that is what is to be done.

Interestingly, you will learn the tricks here that although there are only 2 dimensional data that p cross p, but you are you are basically going in 3 different directions ijk, how this is happening in the data set, that part we will discuss with an example.

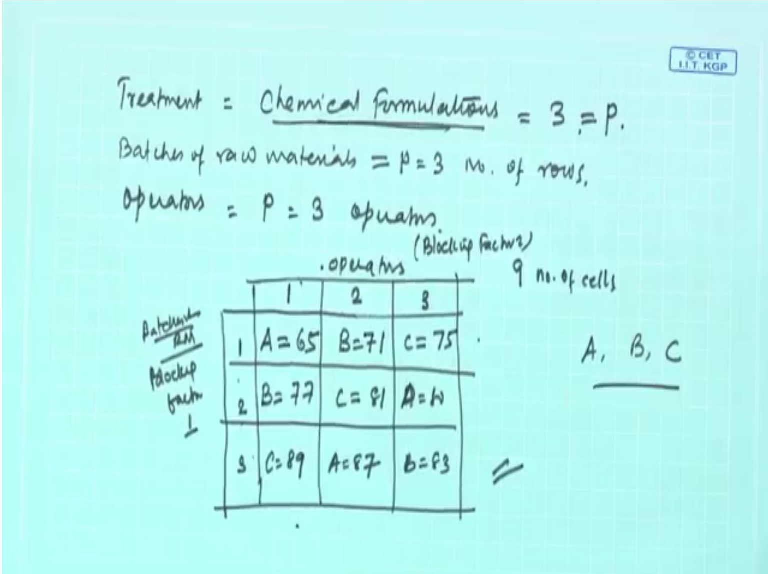
(Refer Slide Time: 18:03)



Here, also we have taken that example for improving chemical yield the process yield, but please remember this data are again assumed data this is not the experimental data. We are assuming that we will be doing experiment and based on experiment you will be getting similar kind of data. How why it is Latin square and how the Latin square will be analysed.

Here, the treatment interested 3 chemical formulations A, B and C. 3 chemical formulations A, B, C; 2 blocking factors are batches of raw materials and operators. We have treatment chemical formulation. How many? 3 in number.

(Refer Slide Time: 19:19)



So, go to p equal to 3, then what is the first blocking factor? batches of raw material. How many batches required here, we require p equal to 3 batches, this is 3 numbers of rows. Then second one is our operators. This is the second visual factor, it also will be having p number of p equal 3, so, 3 operators needed. Then in this case you see we have suppose, this side operator 1, operator 2 and operator 3, this is operator then this side suppose batches of raw material 1, raw material 2 and raw material 3. This is our raw material.

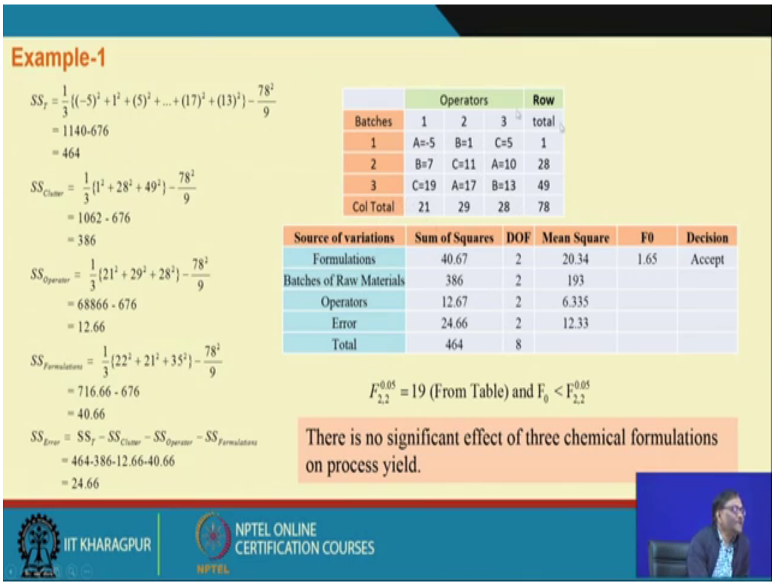
Now, we have that is why 3 cross 3 nines number of cells; cell 1, cell 2, like this. Where is the treatment, this operator is the blocking factor 2 and batches of raw material is blocking factor 1, batches of raw material. Then, each cell you are putting the chemical formulation; chemical formulation is A, B and C, you put in such a manner that the orthogonality is maintained.

Suppose, first cell is A; that means, that chemical formulation A with batch raw material 1 and operator 1 is experimented and the result process yield is therefore, 65. In that case, similarly, B process yield is 71, C process yield is 75, then again here B process yield is 77, C process yield is 81 and then A process yield is 80, then C process is 89, A process yield 87, B process yield 83. Each cell also cell in the cell that treatments are placed in such a manner that you are getting the yield also and that is why, 1, 2 and each cell that the 3 directions are captured here.

Here, you see in each of row there every treatment occurs once, each column every treatment occur 1 and it is a square and it is Latin square. That mean, it is the for the example go to slide for the example what we have shown here it is a fit case for it is a Latin square case. Now, how to analyse the data and get the resultant information what we are seeking for is, whether the chemical for 3 chemical formulations are differently affecting the process yield or there is no difference in the chemical formulations in terms of producing yields.

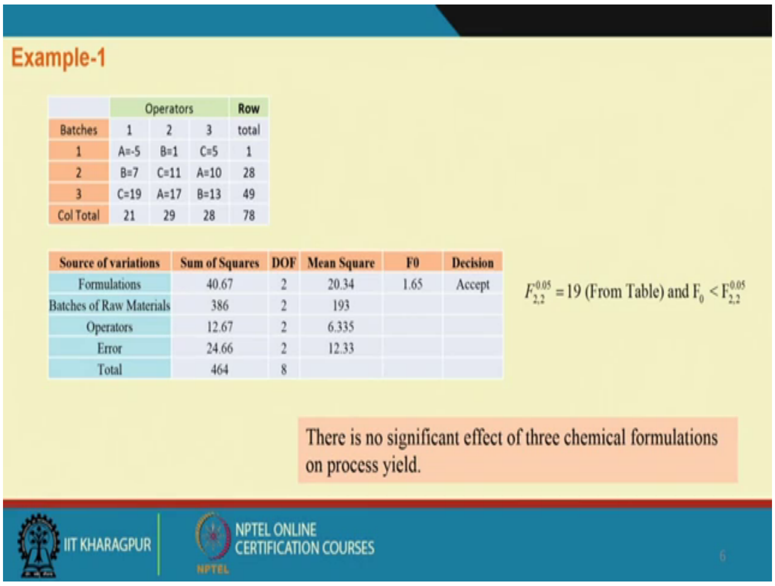
Just to for the sake of algebraic arithmetic simplification, suppose if you subtract each of the observations here by 70, then you will be getting this is the reduced matrix and if you do analysis use the formula what I have already shown you with these data or these data there will no difference because every observation is subtracted by a constant, that will not create any problem in the SS calculation.

(Refer Slide Time: 23:58)



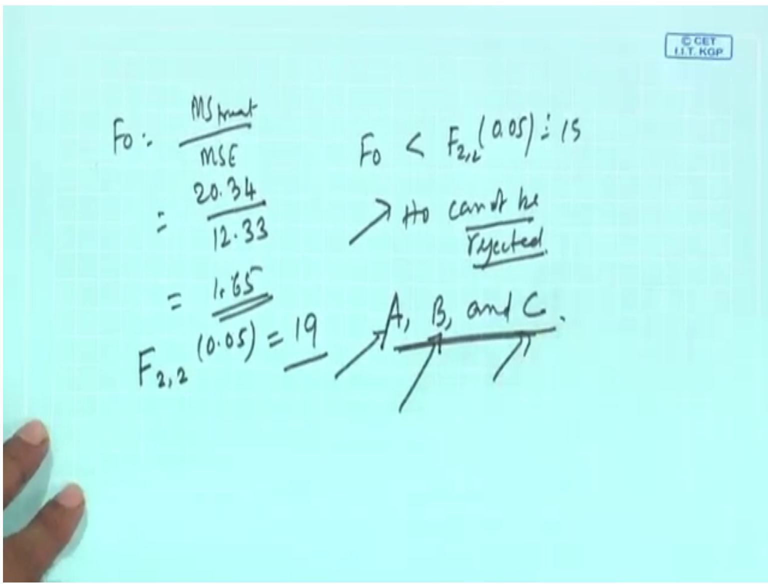
Using this data we have calculated SS T, SS T is what SS T will be what is the formula for SS T? 1 by p, j equal to 1 to p and then y dot j square by N, sorry this is SS treatment. SS T is this 1 sorry SS T is the sum total.

(Refer Slide Time: 24:41)



I think then, with reference to this example, then how do compute the SS T, SS treatment and SS rows, column, SS error these are the formulas will be used. Suppose, if I use this formula for SS T then you see that we will be getting SS T 464 and formula for SS formulation will give this, these are the values for sum square we are getting. And, here we have 3 formulations, p equal to 3, degree of freedom will be 3 minus 1, 2 again 2, 2, 2 and finally, your total is number of observation is 9 minus 1, 8 will be the degree of freedom for total, then calculate the mean square.

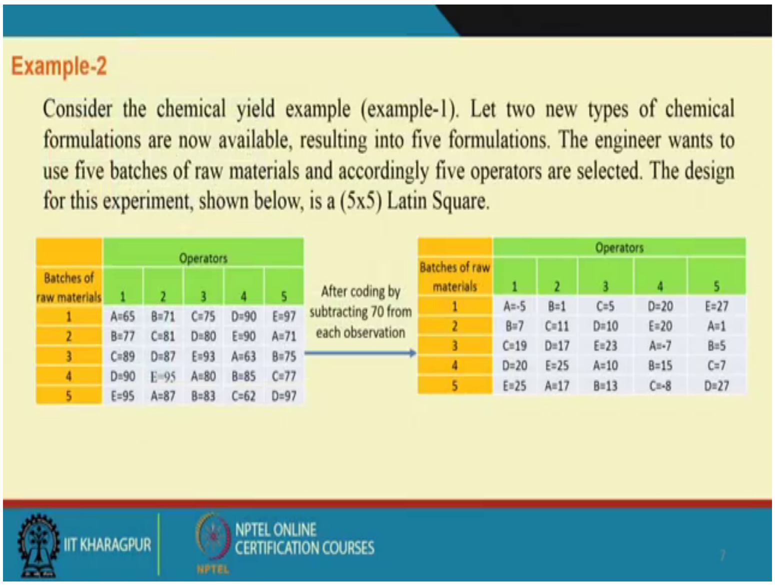
(Refer Slide Time: 25:48)



And, what we are finding out that the for the treatment the F 0 value, what is F 0 is MS treatment by MSE; in this case, MS treatment 20.34 and MSE 12.33 this one giving us value of 1.65 which to be compared with F, p minus 1 p minus 1; 2, 2 let alpha is 0.05 and this value become 19.

As a result, F 0 is less than F 2, 2 that 0.05 which is 19, H 0 cannot be rejected. So, that means, the chemical formulation A, B and C their effect on the process yield is similar or same. There is no significant difference in the mean yield, if you go for chemical formulation A or use chemical formulation B or use chemical formulation C.

(Refer Slide Time: 27:10)

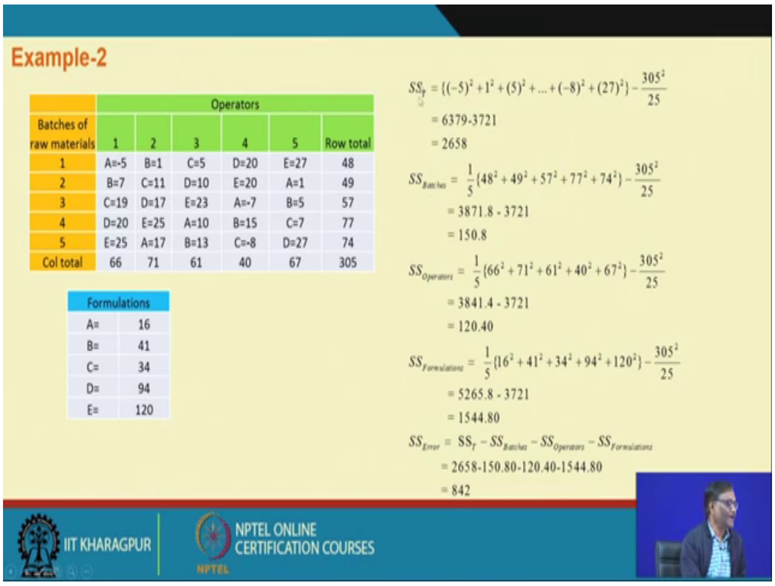


Suppose, there are 2 new chemical formulation available apart from these 3, 2 more new chemical formulations are available in the market and it is told by the form the company that these 2 new chemical formulation giving you better result or their if you use these the position will be better or more than the existing ones.

The engineer in charge he wanted to that do an experiment and to test whether the 2 new chemical formulations are really good or they are also same. Here, what happen, you want to use a Latin square design because of restrictions on the number of experiments to be conducted, as there are 5 chemical formulations, the engineer chosen 5 different batches of raw materials and 5 different operators and the result results for different way use different formulation the yield given here.

Here, also we have subtracted the original observation experimental data each of the cell data by 70, but it can be subtracted by some other value like 100 also, not 100 may be by 90. Then, this value will become little smaller values. It is ok; whatever value you subtract a convenient value will be subtracted so that you will handle data with less amount that the calculation part becomes easier. As in the previous example, I have subtracted by 70, we keep the same one here.

(Refer Slide Time: 29:10)



Now, using this formula SS T, SS batches, operators and formulations and then what happen we will found that SS T is this SS batch is this, operator this, formulation this and SS error this.

When you go for computation of SS batches, that is what I wanted to tell in the first example also, but here let me tell you clearly, what I am saying; when you are talking about SS batches, what is the formula you are using? The formula you are using suppose, SS batches mean SS rows, so, y i double dot square, that mean, row total square. When you are going for operators that column blocking variable that column total square you are using here.

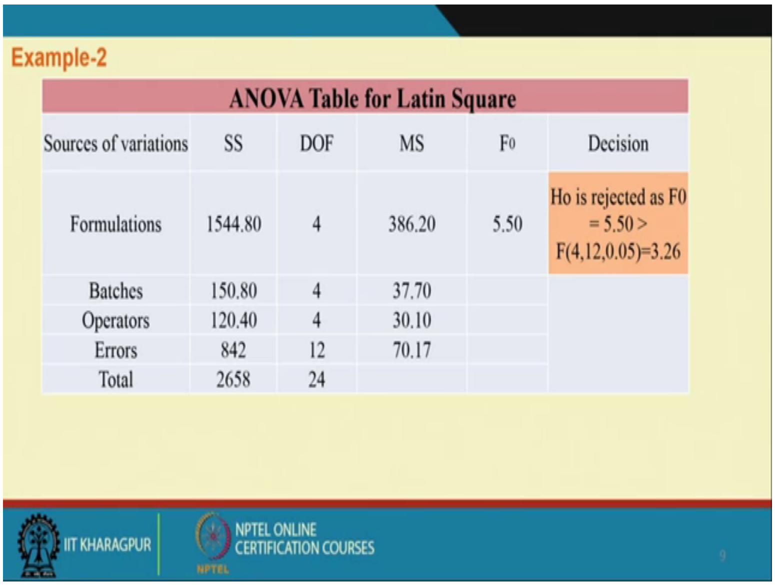
What is row total? First row 48, second row 49, third row 57, like this you see SS batches, all those row total are squared and their sum is taken and then we divided by 5 because there are 5 batches of raw material. We have taken the average of those squared values and then it is subtracted by the correction factor that is y triple dot square by N and you are getting this value.

From computation point of view when you do SS total you basically sum square all the observation 25 observations and take their sum and minus the correction factor, 305 the total square by 25, when you are going for row S square rows, the row total each of the row total will be squared and their average will be taken corrected by this value. When you are going for operator that is the column case SS columns, again the column total will be squared and divided by the number of observations in the column total and then subtracted by the correction factor.

When you go for formulation then, what is that value that is the interesting one. You see here how many formulations A, B, C, D, E there are 5 formulations are there, you take the total for all the A's. First column-1 minus 5, column-2 17, 17 minus 5 is 12 column-3 it is 10, that mean 12 plus 10 is 22, 22 here it is minus 7. So, 22 minus 15 and here it is 1; that means 16. That mean only the values attached with A will be totalled. That is the total formulation that is 16, for A 16.

Similarly, for B users total this one for first column B equal to 7, second column B equal to 1, third column B is equal to 13, fourth column B equal to 15, fifth column B equal to 5 all those things when you total you are getting 41. In that manner, you will get the total for different formulation and here, when you are calculating the formulation you are writing SS formulation equal to 1 by 5, 16 square plus 41 square plus 34 square plus 94 square plus 102 square minus this case and this is the case, then SS error by subtraction.

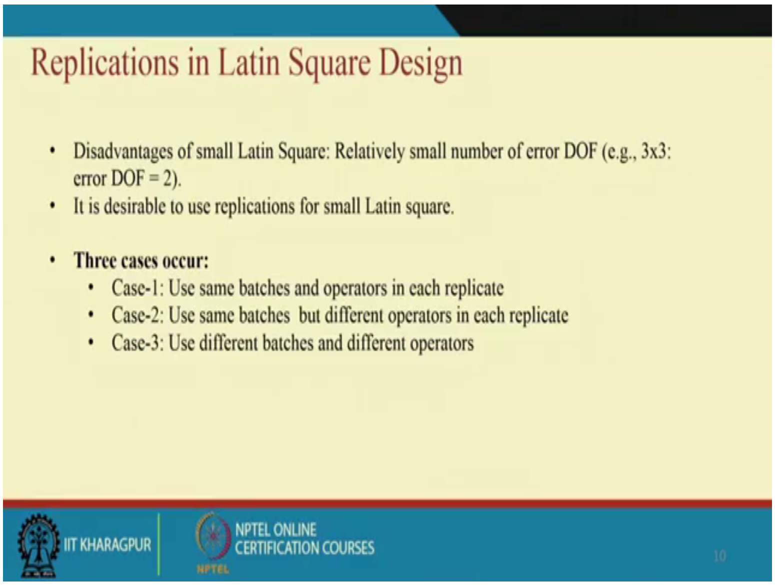
(Refer Slide Time: 32:55)



Sources of variations	SS	DOF	MS	F0	Decision
Formulations	1544.80	4	386.20	5.50	Ho is rejected as F0 = 5.50 > F(4,12,0.05)=3.26
Batches	150.80	4	37.70		
Operators	120.40	4	30.10		
Errors	842	12	70.17		
Total	2658	24			

And, here what happen, when you do the ANOVA test you find out that F 0 computed is 5.50, but here error degree of freedom increases to 12 and MS errors 70.17 and F 0 is 5.050 and the threshold F 0 is 3.26. Computed F 0 is more than the threshold F 0. There is formulation effect. So, that mean I think that that 2 new formulation that came to the market they are giving better results.

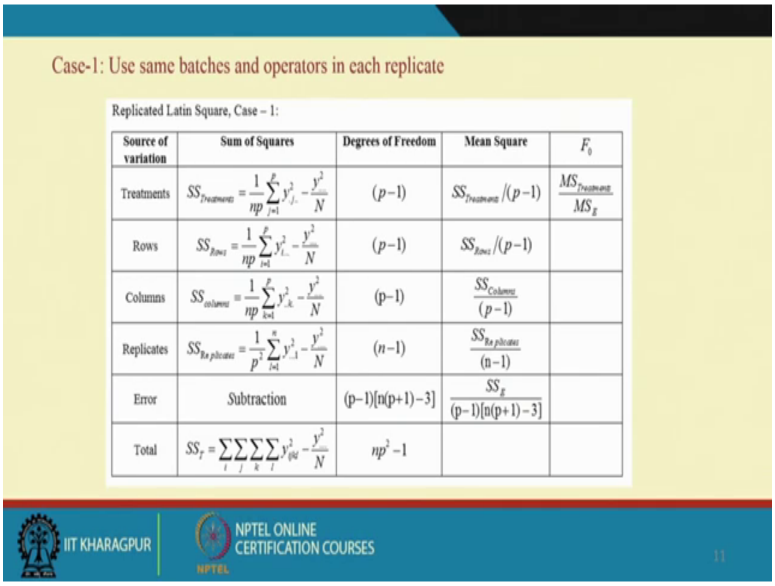
(Refer Slide Time: 33:45)



This is what is our Latin square and there are some more things in Latin square which you required to read by self-study, that is, that when the number of treatment or treatment levels are less like 3 3 or 4 4, there you have found out that the error degree of freedom become very less and it may give you better results. As a result, you require using replications in case of Latin squares and particularly in the small Latin squares.

There will be 3 cases; case 1, use same batches and operators with reference to the example, use same batches and operators in each replicate. You do replication, but keep by batch and operator same. Other case is use same batches, but different operators, one batch different replicates with different operators and then case 3, use different batches and different operators.

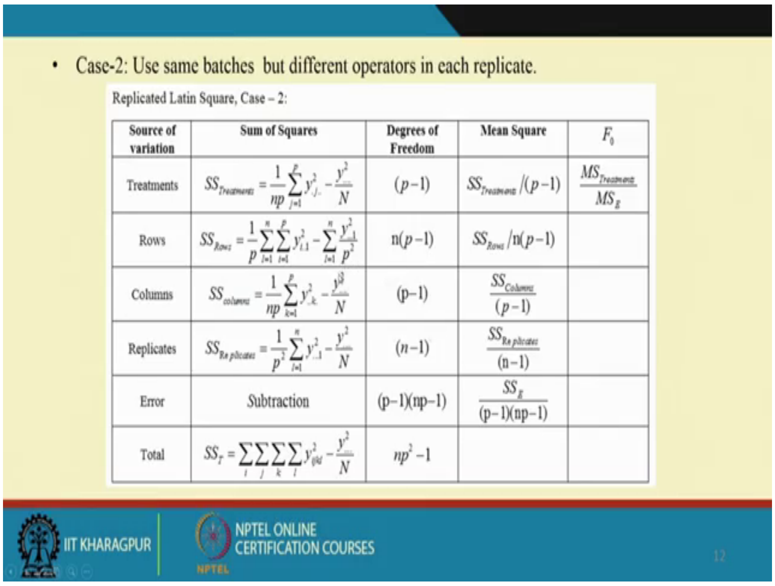
(Refer Slide Time: 34:45)



Source of variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{Treatments} = \frac{1}{np}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}$	$(p-1)$	$SS_{Treatments}/(p-1)$	$\frac{MS_{Treatments}}{MS_E}$
Rows	$SS_{Rows} = \frac{1}{np}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{N}$	$(p-1)$	$SS_{Rows}/(p-1)$	
Columns	$SS_{columns} = \frac{1}{np}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}$	$(p-1)$	$\frac{SS_{Columns}}{(p-1)}$	
Replicates	$SS_{Replicates} = \frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}$	$(n-1)$	$\frac{SS_{Replicates}}{(n-1)}$	
Error	Subtraction	$(p-1)[n(p+1)-3]$	$\frac{SS_E}{(p-1)[n(p+1)-3]}$	
Total	$SS_T = \sum_i\sum_j\sum_k\sum_l y_{ijkl}^2 - \frac{y_{....}^2}{N}$	np^2-1		

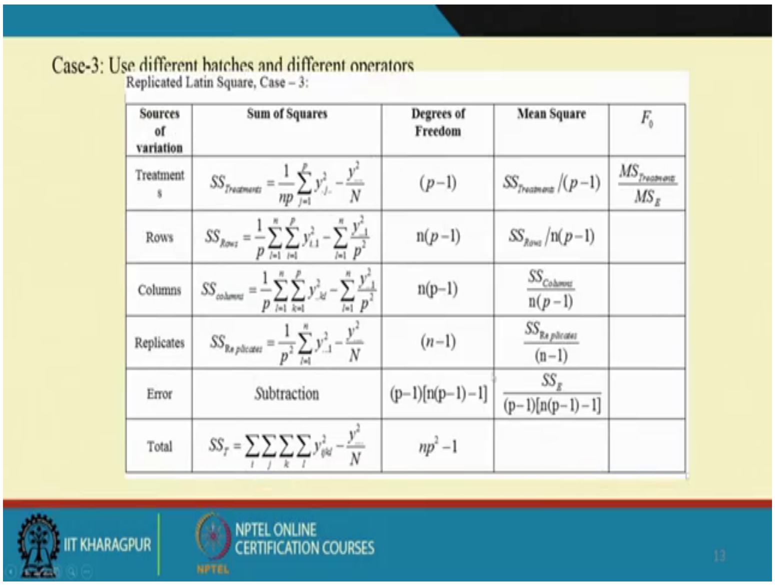
What will happen accordingly, the formula will change the resultant tables will be like this; this one.

(Refer Slide Time: 34:47)



This is for case – 2.

(Refer Slide Time: 34:49)



This is for case – 3. These are available in Montgomery book.

And, this is what is our Latin square design and I hope that a simple that what is Latin square and when it will be used, it is used for blocking simultaneously blocking 2 nuisance factor and with 1 treatment and hope that you will be able to reproduce the results and you have you got the concepts also.

Thank you very much.