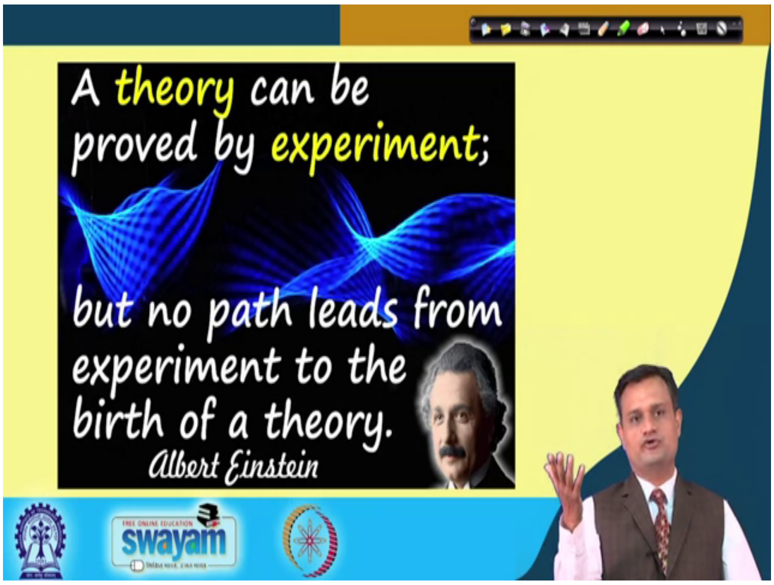
Six Sigma
Prof. Jitesh J. Thakkar
Department of Industrial and System Engineering
Indian Institute of Technology, Kharagpur

Lecture - 39
Randomized Block Design

Hello friends, you are welcome to our ongoing Six Sigma journey and we are in the improve phase of our DMAIC cycle. So, we have opened our discussion on design of experiment as a part of improve phase of DMAIC cycle and we had seen the basic concepts introduction of design of experiment in the last lecture. So, now, as a part of lecture 39, we will talk about Randomized Block Design.

(Refer Slide Time: 00:53)

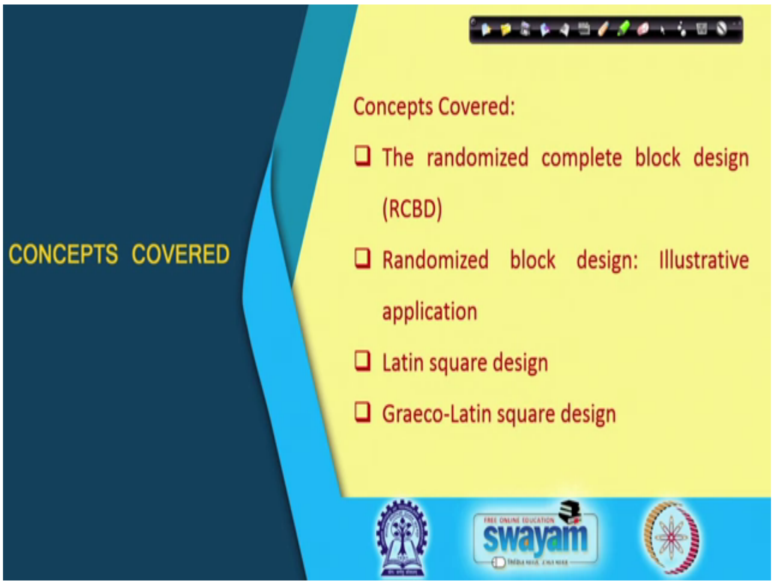


Once again I would like to remind you that a theory can be proved by experiment and no paths leads from the experiment to the birth of a theory. So, Albert Einstein firmly believes that, whatever theory you are proposing it must be proved by a scientific procedure of experimentation.

(Refer Slide Time: 01:15)

If I just give you the recap we talked about the history of the DOE, the contribution of fisher Taguchi another say great mathematicians; DOE principles and terminologies, the key experimentation strategies then guidelines for planning conducting, some of the key terms we discuss there is a factor level treatment, noise factor, controllable factor, uncontrollable factor and I would once again mention that in order to conduct effective design of experiment, you must have a domain knowledge.

(Refer Slide Time: 01:54)

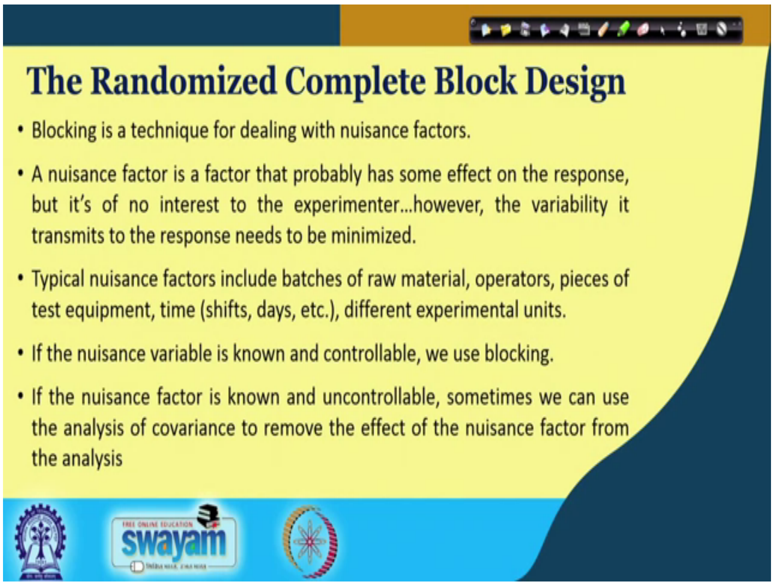


So, this particular lecture, we will talk about randomized complete block design that is

RCBD and we have seen the concept of blocking. So, just to remind you there are couple of terms we have seen 1 is the repetition, other is the replication and third is the blocking. So, this terms you should keep in mind repetition and replication invariably we will use as a part of our design of experimentation and blocking is a case where I am trying to block a particular factor, which has an impact on my results of the experimentation. It may be a shift to shift variation it may be that batch to batch, it may be that I am purchasing the raw material from different vendors and that has some impact on the response variable.

So, randomized block design we will see the illustrative application and then there are some extensions improvements rather that is Latin square design in Graeco-Latin square design as a part of your RCBD that is Randomized Complete Block Design.

(Refer Slide Time: 03:08)

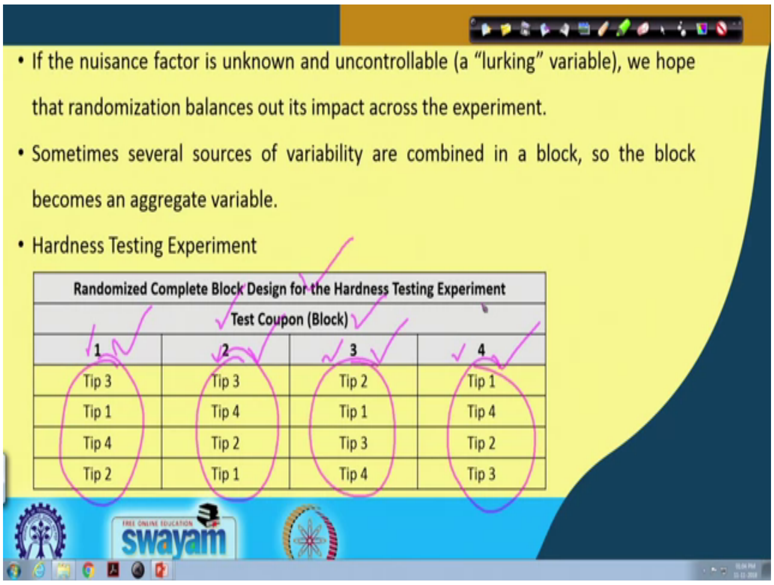


So, what is randomized complete block design? So, blocking as I mentioned is a technique for dealing with the nuisance factor and you may have raw material you have operator shift to shift variation time test equipment as the blocking as the factors to be block and the nuisance variable is known and controllable.

Now, there is something interesting to appreciate that, you have the factor which is controllable and known. So, you know that yes vendor one, vendor 2 and vendor 3 there is difference in the quality. So, you know, but you are not able to control. So, this is your nuisance factor which is controllable and known.

Now, let us say if you have your nuisance factor, which is known and uncontrollable sometimes we can use the analysis of covariance to remove the effect of the nuisance factor from the analysis. So, note that we apply the blocking for all. So, just appreciate this 2 point if I have controllable and known nuisance factor then only I can conduct RCBD that is randomized complete block design I cannot block the factor if it is not if it is not controllable or uncontrollable, then I will go for some covariance kind of strategy to minimize its impact.

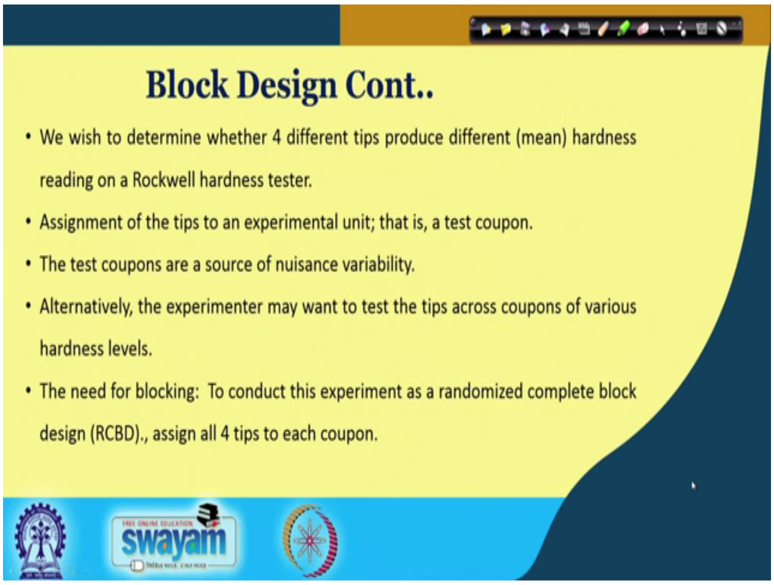
(Refer Slide Time: 04:42)



Now, just see the example here and what is the example that, you have some uncontrollable factor lurking variable also it is called and let us say I have a machine for testing the hardness of the specimens or the material and I have some test coupon. So, you can see here that randomized complete block design for the hardness testing experiment there are test coupon, which are considered as the blocks.

So, I have 1 2 3 and 4 as the test coupon and I am using different tips for my hardness testing process and this tips are randomly say assigned in a particular subgroup or particular block rather you say this is block 1 this is block 2, 3 and 4. So, you have the test coupons as the block and you have within the block different tips to be used for having the hardness reading.

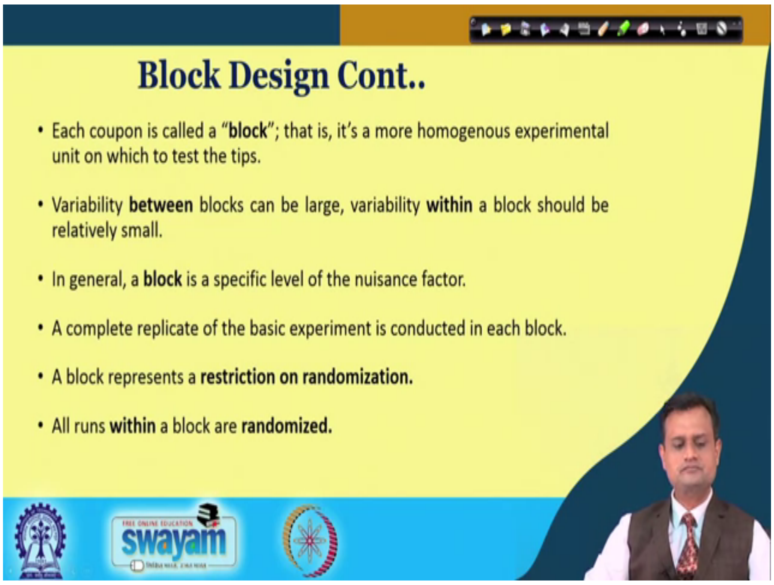
(Refer Slide Time: 05:55)



So, this is my example of a blocking for hardness testing, now as I mentioned we was we wish to determine whether 4 different tips produce different mean hardness reading on a Rockwell hardness tester. So, I am using different tips if you have manufacturing knowledge, now whether this tips will produce the same hardness reading or there is a difference.

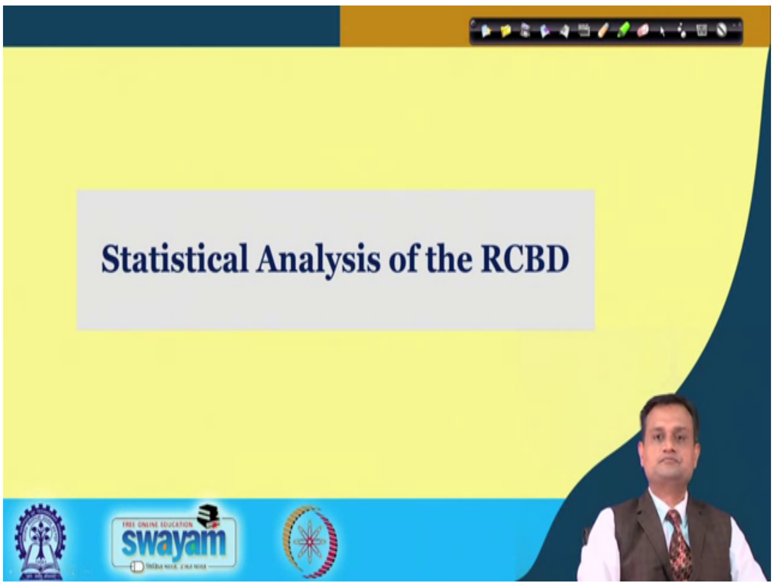
The nuisance factor here is typically, the my test coupon an assignment of tips to an experimental unit is my test coupon. So, I would like to block particular test coupon and then for that particular test coupon I will have the readings of hardness reading with different tips.

(Refer Slide Time: 06:46)



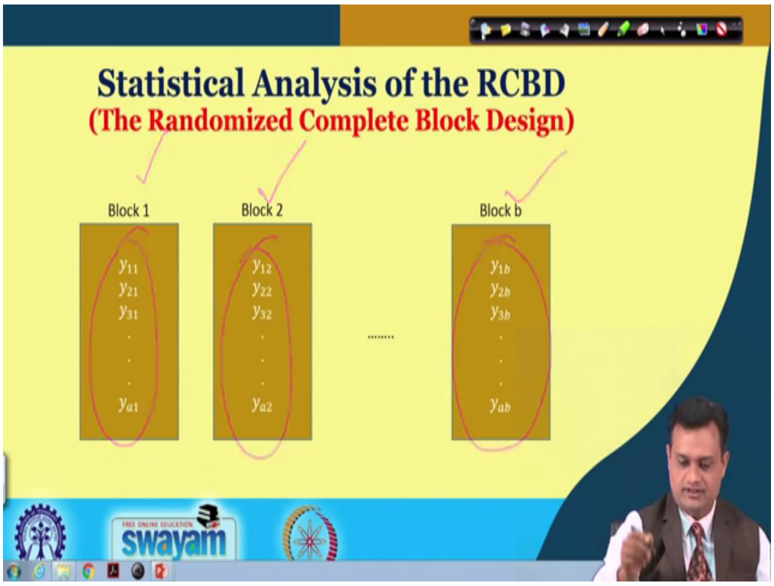
So, we can conduct this through RCBD. So, each coupon is called block and it is more homogeneous experiment unit on which the test the tips. Variability between so, now your ANOVA knowledge will come in use and block can be large variability within a block should be relatively small that is my understanding and a block is specified level of the nuisance factor. So, in this case your block or nuisance factor is the test coupon. So, now a block typically represents restriction on randomization and all runs within the blocks are randomized.

(Refer Slide Time: 07:36)



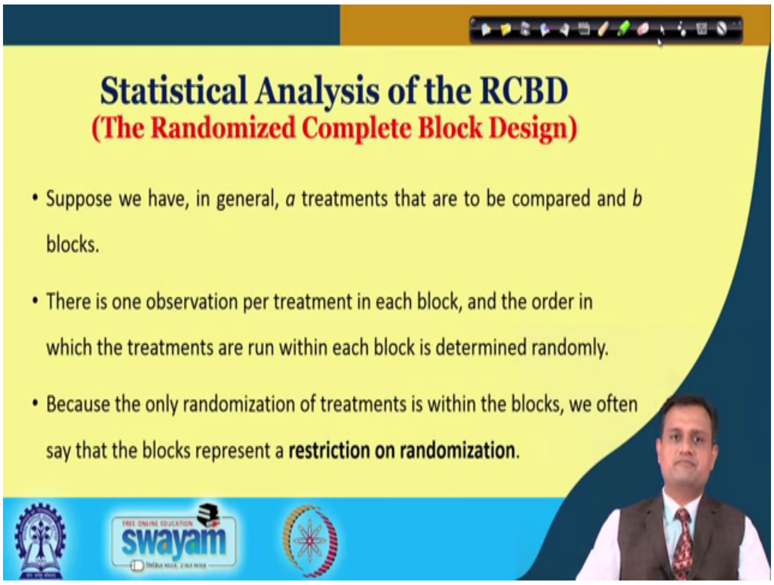
So, allocation of the tips within a particular test coupon is randomized and this is where say I am trying to use the concept of RCBD. Now statistical analysis of RCBD goes like this.

(Refer Slide Time: 07:42)



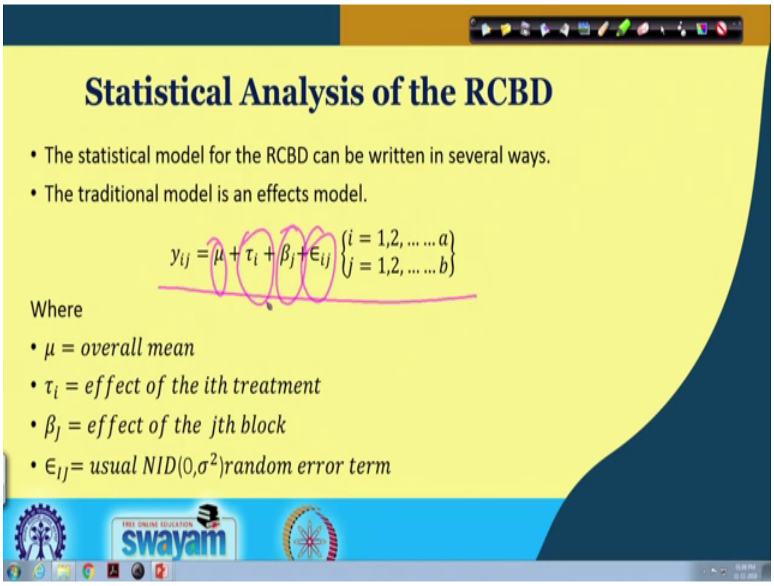
You have the block 1, you have the block 2, you have the block b and you have the various readings within this that is the outcome of your reading of your response variable.

(Refer Slide Time: 08:00)



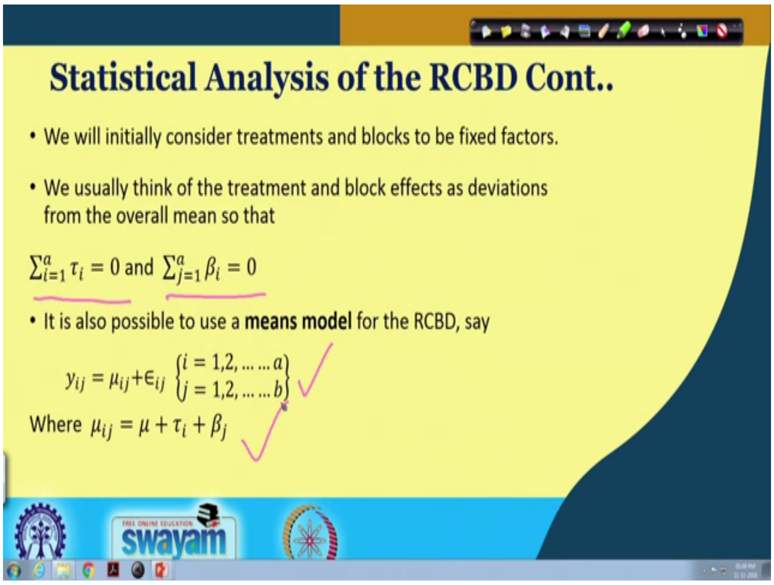
Now, suppose we have in general a treatments, that are to be compared and b blocks. So, each particular say column you can say that it is a block and then you have a treatments to be analyse in row. So, there is an observation per treatment in each block and the order in which the treatments are run within a block is randomly as I mention. So, because you have the randomization only within the block, your RCBD put some restriction on the randomisation that we accept.

(Refer Slide Time: 08:40)



Now, just appreciate the model that you have y ij and basically y ij is your response dependent variable which is mu overall mean tau i that is the treatment effect beta j effect of the jth block because you have created the blocks shift 1 shift 2 shift 3 vendor 1 vendor 2 vendor 3 or here it is the test coupon 1, test coupon 2, test coupon 3 and so on and then your model will have some say variation which we assume to be normally distributed.

(Refer Slide Time: 09:30)



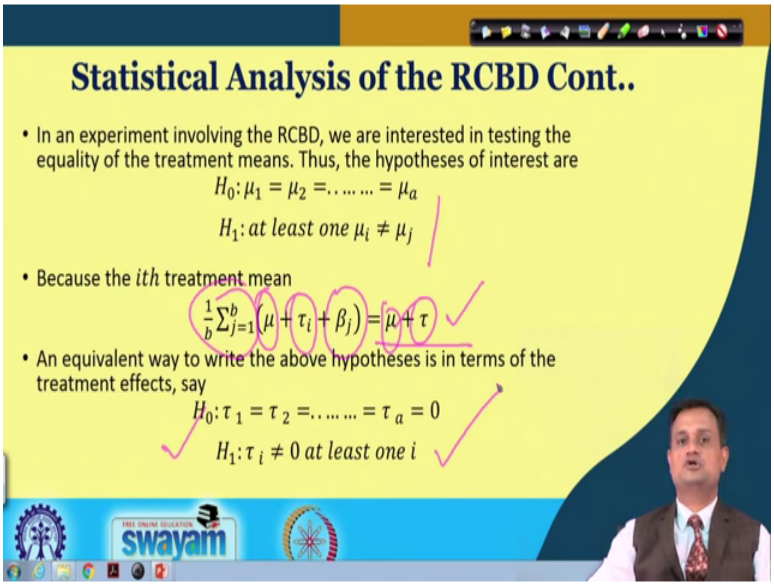
So, now with this basic understanding of my statistical model let us try to say go ahead with some understanding that if you think that treatment and block effects as deviation from overall mean then we can set this to be like this that sigma is equal to 1 to a tau i is equal to 0 it means there is no treatment effect and then there is no block effect is equal to 0.

So, I may set my null hypothesis and alternate hypothesis to investigate, that whether there is a block effect or there is a treatment effect and say yij is mu ij plus epsilon ij and this is my y mu ij is mu plus tau i plus beta j. Just to make understanding very simple, let me help you that, suppose you are purchasing the material from different vendors and this vendors are basically given the feedback, that the quality of your raw material is not adequate.

Obviously, this vendor may say that our raw material is but the kind of processes you have machines you have, they are having larger variability producing larger variability and hence there is no problem with our material, how to resolve such issue? And that to vendors know that you are purchasing the raw material from different vendors. So, vendor may claim or may enter into the blame game, that no its not my problem it may be with the problem with the vendor 2, may be problem with the vendor 3 and you are mixing the material raw material received from different vendors; so, now, your lost.

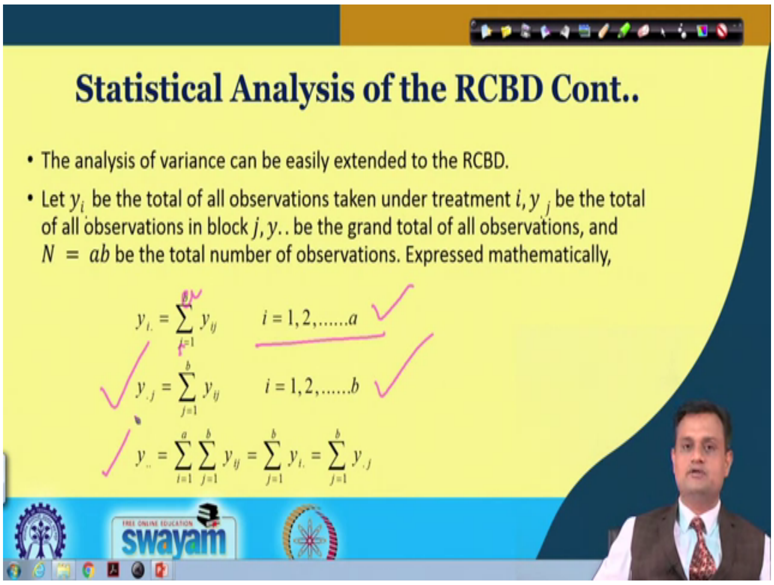
So, in this case you can conduct the experimentation by blocking, the material received from each vendor and then you conduct the experimentation to see that whether the block effect is significant or not or there is mainly because of machines or my processes, and this is how you can come to a statistically sound inferential conclusion which can help you to better understand whether all the 3 vendors are different in terms of quality or at least one of them and which one is different. So, this is the way you can resolve many conflicts and negotiate better in your business situation.

(Refer Slide Time: 11:58)



So, my null hypothesis would like this, we cannot go away from this because its an inferential statistics and I would say that mu 1 is equal to mu 2 is equal to mu a. So, I am considering the treatment a refers to the treatment, I will say that my 1 by b sigma j is equal to 1 to a mu plus tau i plus beta j is equal to overall mean plus treatment then I have H 0 that is my treatment. So, I will say tau i is equal to tau 2 is equal to tau a that is my various treatments in the row and this is true if one of the treatment differs.

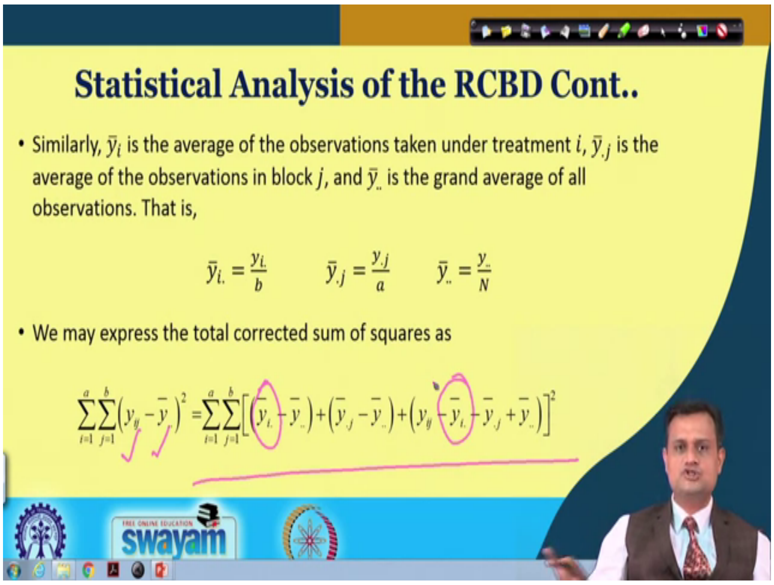
(Refer Slide Time: 12:44)



Now, let us try to see some basic equations which can easily be interpreted when I say y i dot. So, once again I would like to remind you that I am referring the book Douglas Montgomery and the notations are used as per this text only if you refer some other text the basic procedure will remain same, but there may be some change in say defining the various variables or defining the parameters so, that you need to adopt.

So, I am using the notations from Douglas Montgomery and y i dot is equal to sigma j is equal to 1 to b that is your column y ij and j is equal to 1 to b this is your 1 to b or here you can just change it this is y i. So, I will just make little correction this is my 1 to a and this is i is equal to 1 to a. So, this would be my i; this is my y dot j. So, j is equal to 1 to b yij i is equal to 1 to 2 to b, this is my block, this is my individual treatment y double dot is i is equal to 1 to a j is equal to 1 to b yij and I can put it like this. So, this is a very simple thing to get the summation of the respective quantity.

(Refer Slide Time: 14:15)

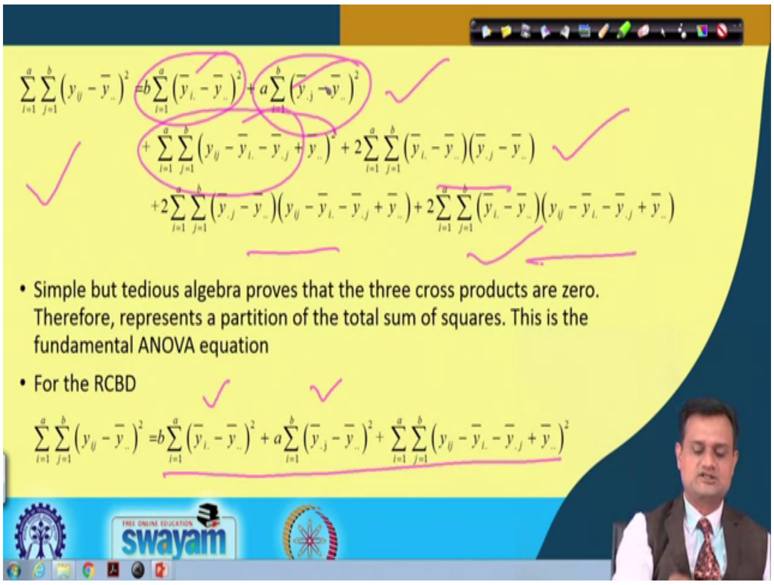


Now, yi dot bar I am taking the b is basically your y i dot divided by b you have found this summation for y i dot similar y dot j bar. So, please be comfortable in speaking the notations y dot j bar is equal to y dot j that is the summation divided by a and this is y double dot bar is y double dot divided by capital N number of readings in your this thing.

Now, let us say I want to express this corrected sum of square when we conduct the ANOVA analysis we deal with the sum of square so, that we can convert it into mean sum of square and find the value. So, I am just taking f y ij minus y bar, that is y ij and y double dot bar. So, this difference will basically give me the error component sum of square I want to analyse and then I can just do little manipulation to expand this particular expression.

So, you can see here that I have included some term which is not there here. So, yi dot then I am subtracting this y i dot and likewise I am just trying to say make it simple so, that I can reach to the particular expression in terms of sum of square. So, with this we can just do little bit analysis with the basic equation.

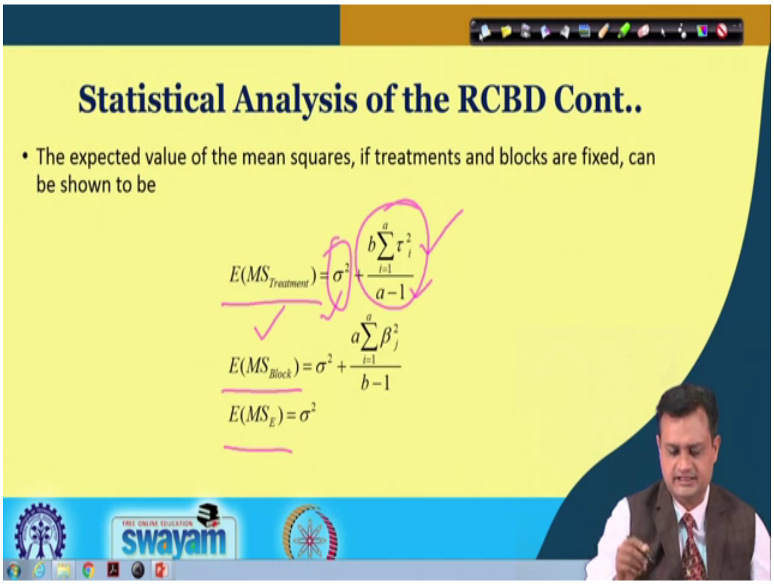
(Refer Slide Time: 15:53)



And what you can see here that, when I am expanding this I would be expanding all the terms. So, you have a minus b whole square you can easily expand it and then its a tedious algebraic process to prove that this cross products are zero, but I am just assuming that some higher order interactions or cross products they are zero and then I am just checking the final version by deleting by crossing some of the say products as zero and I am finally, taking this particular expression. So, this includes basically this term you can see very well this is this term, this is this term and then you have sigma i is equal to 1 to a j is equal to 1 to b.

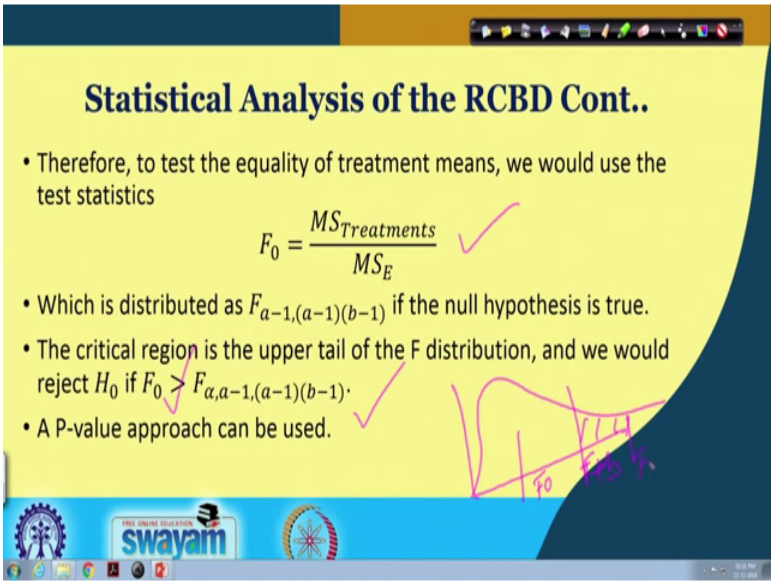
So, you have basically this particular term third term other terms where cross product is there they are set to zero, we are not going into algebraic calculation for proving it to zero, but just they tend to zero accept it and we are just reducing this expression to this.

(Refer Slide Time: 17:08)



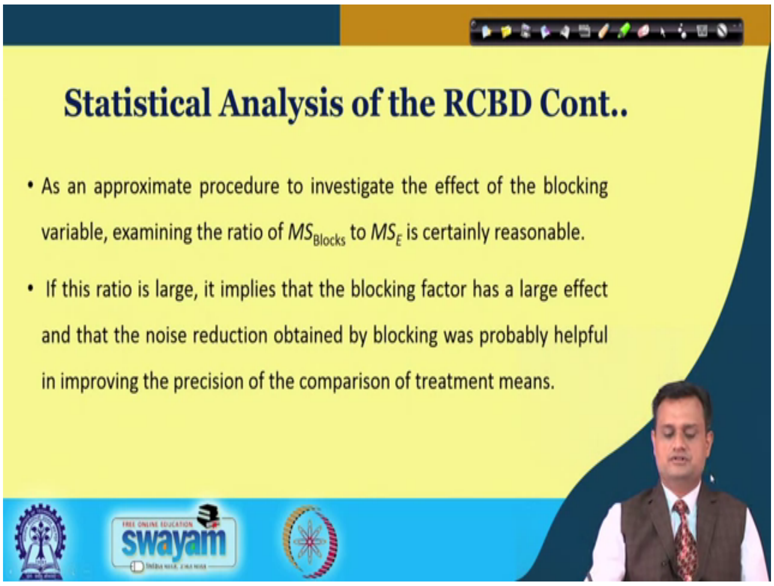
So, you can express your expected mean square treatment as sigma square plus sigma i is equal to 1 to a b in to sigma is equal to 1 to a tau i square divided by a minus 1. So, basically this is the sum of square and when you say consider this treatment effect with the sigma square, that is your population variance you can basically estimate the mean square sum of square for the treatment same block and this is mean sum of square for the error. So, this is what we do in analysing the statistical model DOE model.

(Refer Slide Time: 17:53)

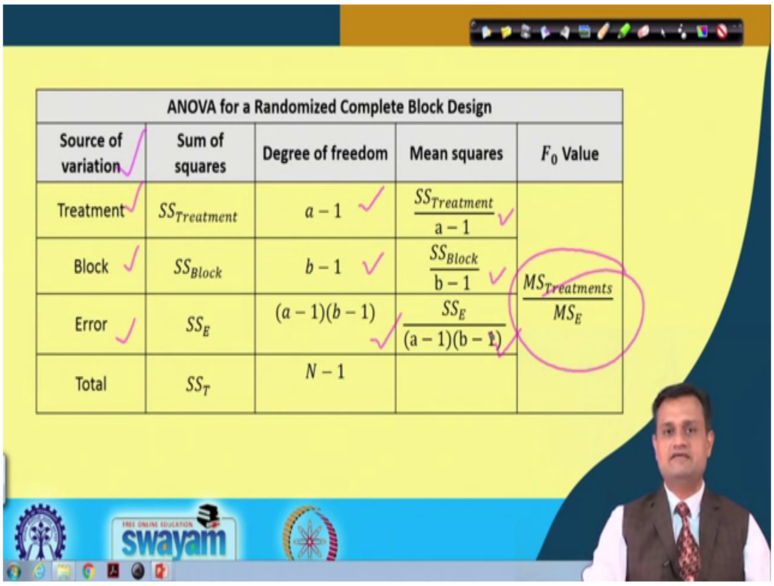


The rule is very simple if you have F 0 that is the calculated value MS treatment divided by MS E and if F 0 is greater than this, then you reject null hypothesis because this will fall into the rejection region. So, basically you have the F distribution and let us say this is your critical value or let us say this is your tabulated value, suppose your observed value falls here then you accept if it falls in this region then you reject. So, this is your observed value. So, basically we try to conduct the analysis same way.

(Refer Slide Time: 18:40)

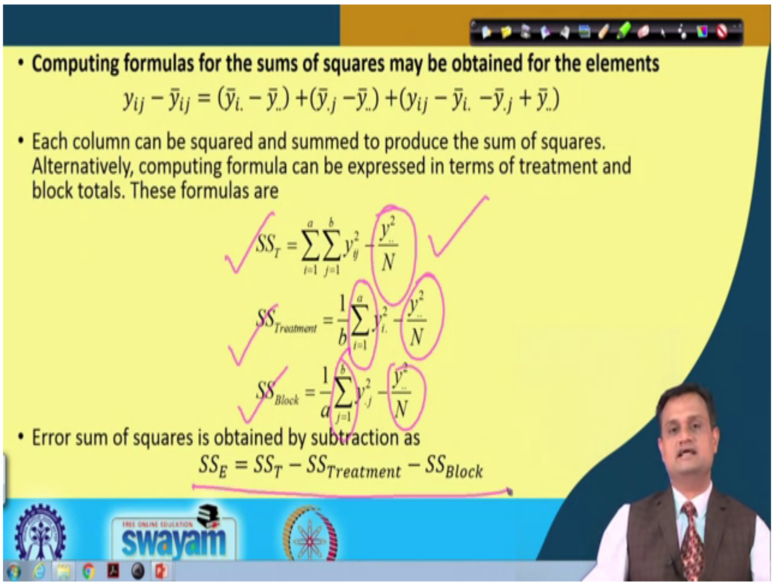


(Refer Slide Time: 18:42)



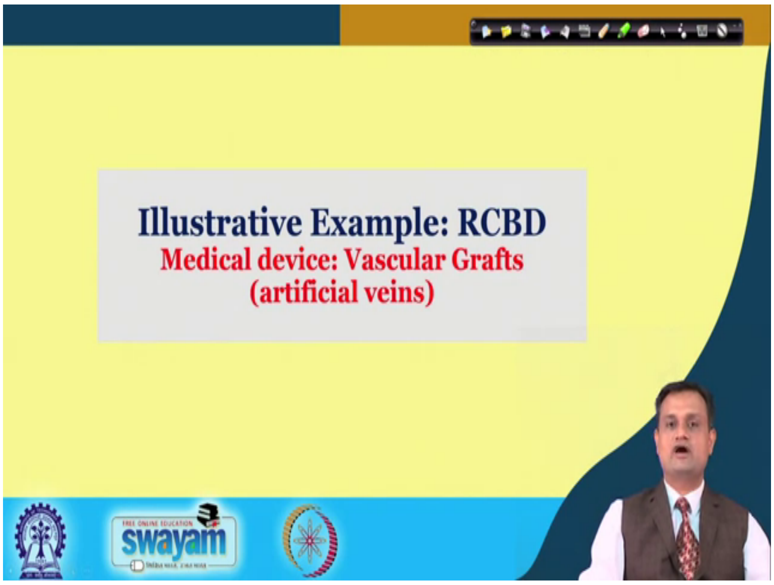
Now, you have this ANOVA table for computing the F statistics treatment value and basically you have sources of variation treatment block error degree of freedom and you have mean square and then you compute the say F value for the treatment to see that when I have blocked a nuisance factor to what extent my treatment has significance some impact on my response variable.

(Refer Slide Time: 19:14)



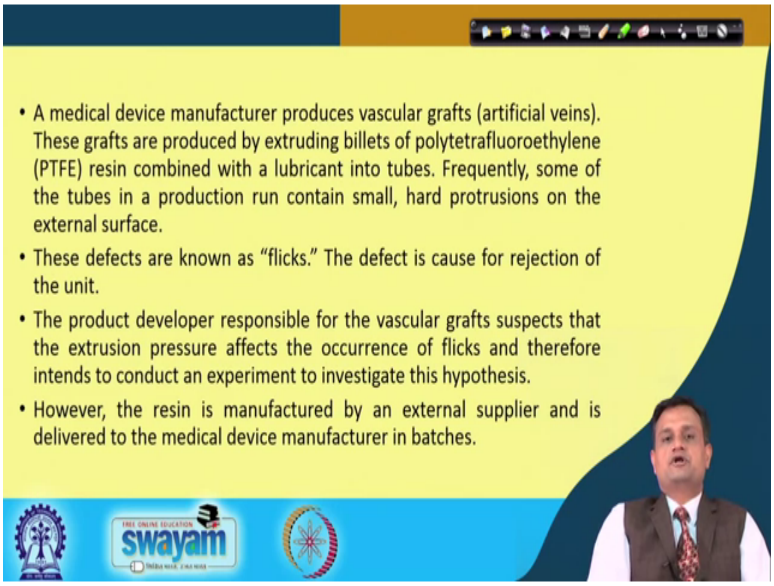
So, this is exactly what we try to do. Now, we need some expressions to compute SST SS treatment SS block. So, this is what you can use sigma is equal to 1 to a j is equal to 1 to b y ij because in a particular cell you can define the reading with respect to i and j i is your treatment j is your block. So, then you subtract this from y double dot square divided by N, this is common in all the equation SST SS treatment SS block because this a treatment you change it from 1 to a row and this is 1 to b say column block. So, this is how and then you compute the SS E by subtracting SS treatment and SS block from the SS total. So, this is how you can easily compute the sum of square.

(Refer Slide Time: 20:05)



Now, let us see an example. So, there is a company they basically manufacture the artificial veins used for the medical purpose and typically called as vascular grafts.

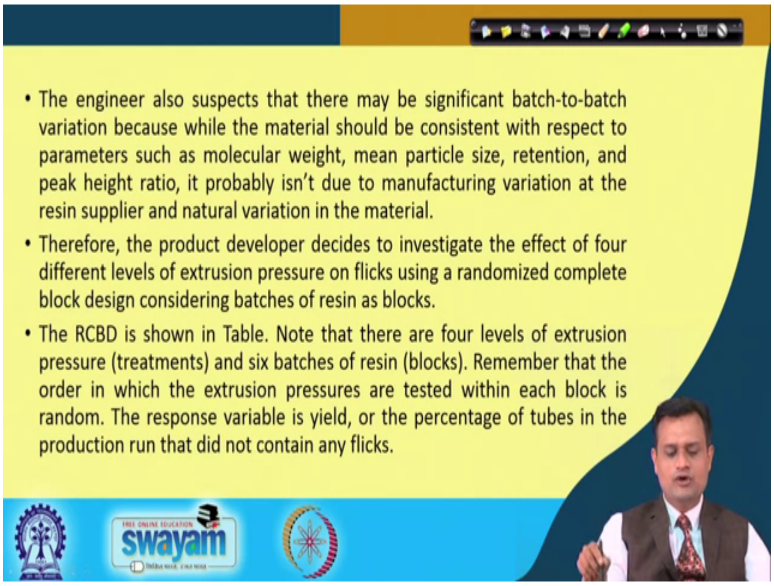
(Refer Slide Time: 20:22)



Now, the quality manager of this particular company as observed, that there is some projected portion there are some flicks on the tubes produced. And then he is trying to investigate that what could be the reason. So, maybe he feels that the pressure which they are applying to produce this particular say tube may be one of the factor or maybe the responsible factor.

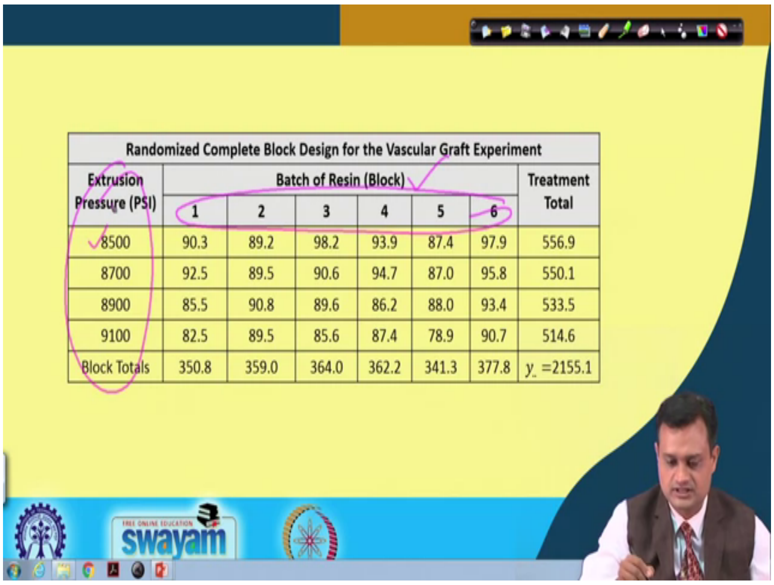
And they do not have in fact, the liberty to purchase the entire resin material for manufacturing this tube from one supplier. So, even they have different block of resin material. So, they want to first block this particular factor that is the resin material. So, that if it is have it has some impact, then this nuisance factor can be restricted to a particular block.

(Refer Slide Time: 21:21)



So, now you have basically the situation like this that you have the block of resins.

(Refer Slide Time: 21:25)

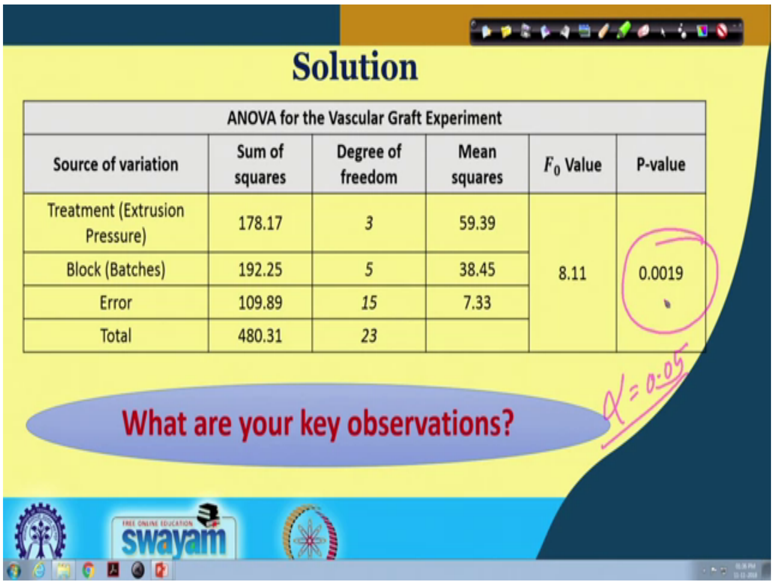


Randomized Complete Block Design for the Vascular Graft Experiment

Extrusion Pressure (PSI)	Batch of Resin (Block)						Treatment Total
	1	2	3	4	5	6	
8500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8700	92.5	89.5	90.6	94.7	87.0	95.8	550.1
8900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9100	82.5	89.5	85.6	87.4	78.9	90.7	514.6
Block Totals	350.8	359.0	364.0	362.2	341.3	377.8	$y_{..}$ =2155.1

So, this resins are basically you have considered 6 different you may consider purchase from 6 different vendors or same vendor giving six different batches of the resins, and here you have considered the treatment that is extrusion pressure. So, now, my interest is to analyze that, to what extent this extrusion pressure is responsible for having flicks or projected portion on the tubes vein tubes.

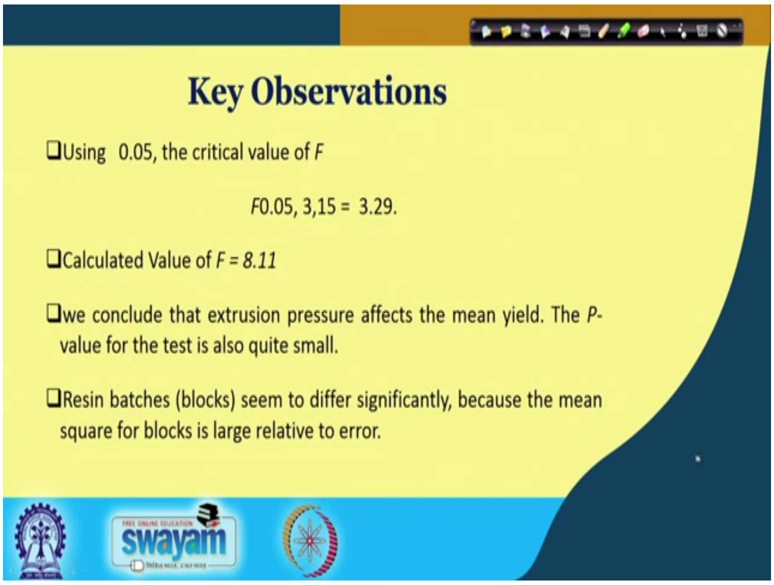
(Refer Slide Time: 22:07)



So, now for this I can easily conduct the ANOVA analysis and this is the ANOVA analysis. So, directly look at the P value. Suppose you are working with alpha is equal to 0.05, then this falls into the rejection region and this clearly says that my null hypothesis about no variation, no impact null is always about statistics in the pressure extrusion pressure that is the treatment is rejected. So, I will say that yes there is an impact of the extrusion pressure on the flicks or projected part observed on the vein tubes.

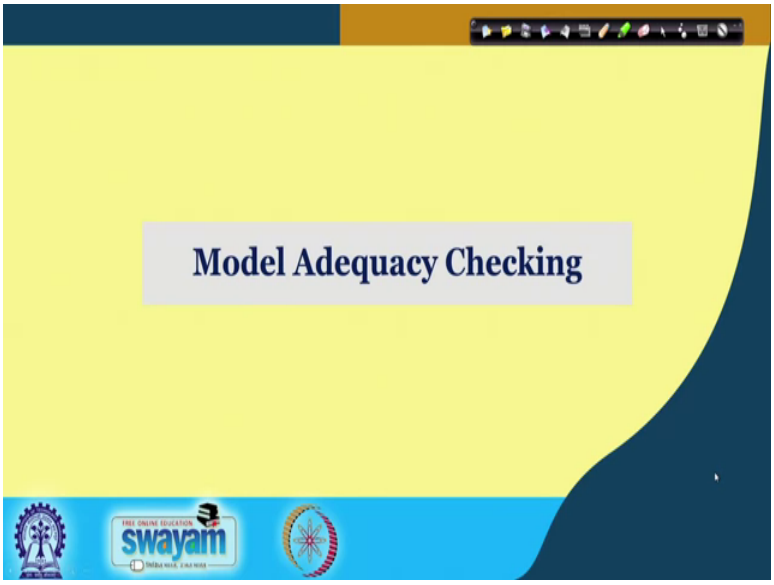
So, once this is investigated, then you would like to do further research and see that what could be the right pressure which can minimize such undesirable effect in terms of flicks on the tube.

(Refer Slide Time: 23:04)



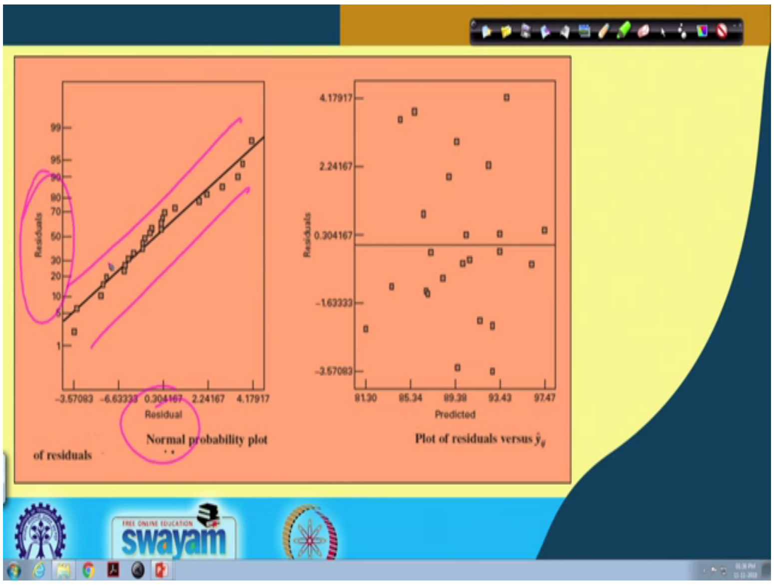
So, this is exactly what I have put here.

(Refer Slide Time: 23:06)



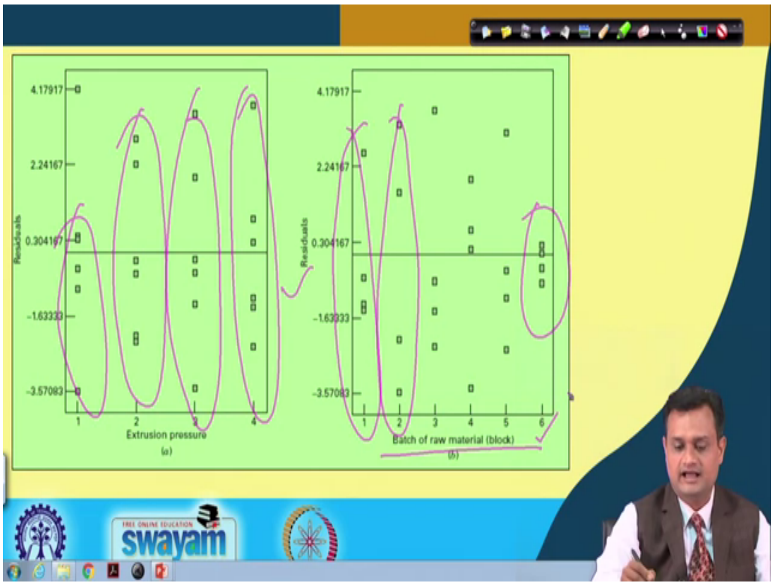
Now, as usual you cannot accept the results without having the model adequacy check.

(Refer Slide Time: 23:11)



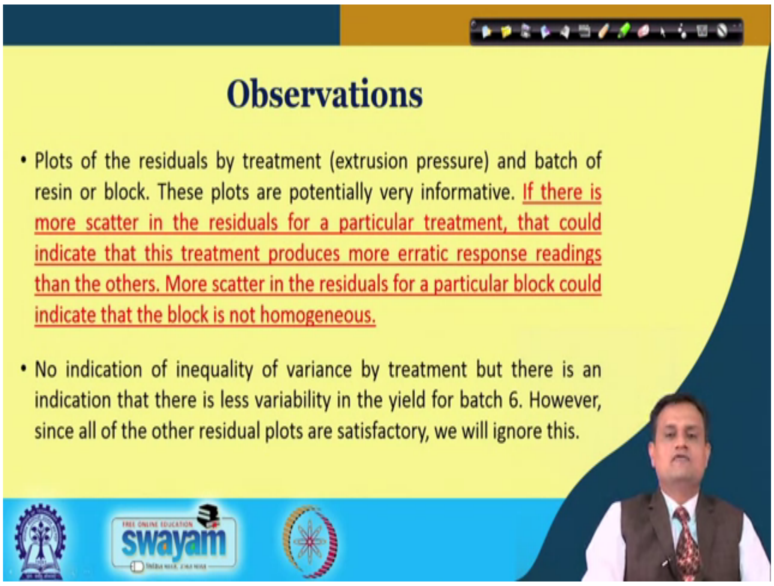
You check the normal probability here it justifies that when I plot the residuals and this is the values taken this is the normal probability plot. So, you plot the various probability value on these and you see that, more or less it is passing through a line. So, my assumption about normality is true. I am also plotting the predicted value and residual I do not see much say trained or pattern in this and this scatter net is justified.

(Refer Slide Time: 23:52)



Now, here if you see the another figure that is extrusion pressure and residual. So, for a given extrusion pressure and residual I can see that each particular say extrusion pressure you have some randomness significant variability captured. Now if you see here this is; this is here it is batch of raw material versus residual. But if you see here then the variability is very very less it means there is some closeness in the residual when I am operating it at 6 number raw material block and this is something that would be of concern.

(Refer Slide Time: 24:38)

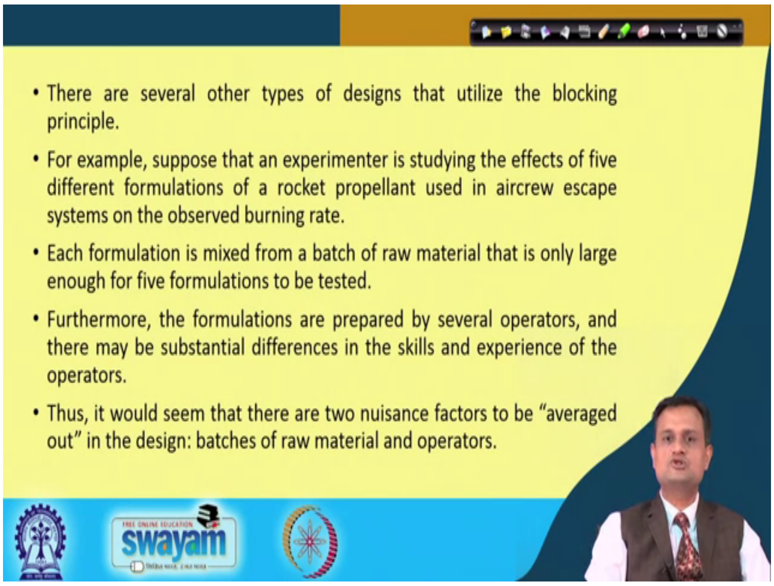


So, we can just say by looking to this 2 figures, if there is more scatter in the residual for a particular treatment that could indicate that this treatment produces more erratic response reading the another and more scatter in the residual for particular block could indicate that block is not homogeneous. Similar way as I mentioned there is an indication there is less variability, but such kind of observations if there is no serious consequence then it can be ignored.

(Refer Slide Time: 25:10)

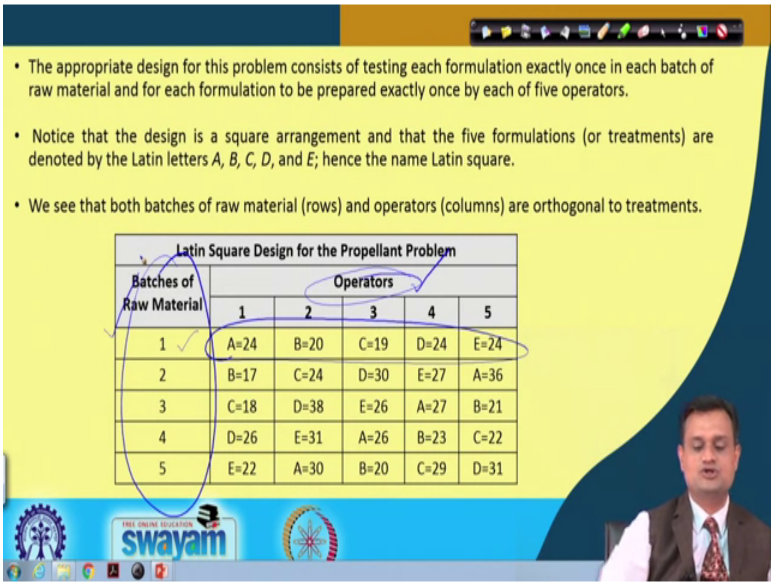
Now, let us see some improvement in the basic RCBD and we have Latin square design its an improvement of the basic design that is RCBD.

(Refer Slide Time: 25:20)



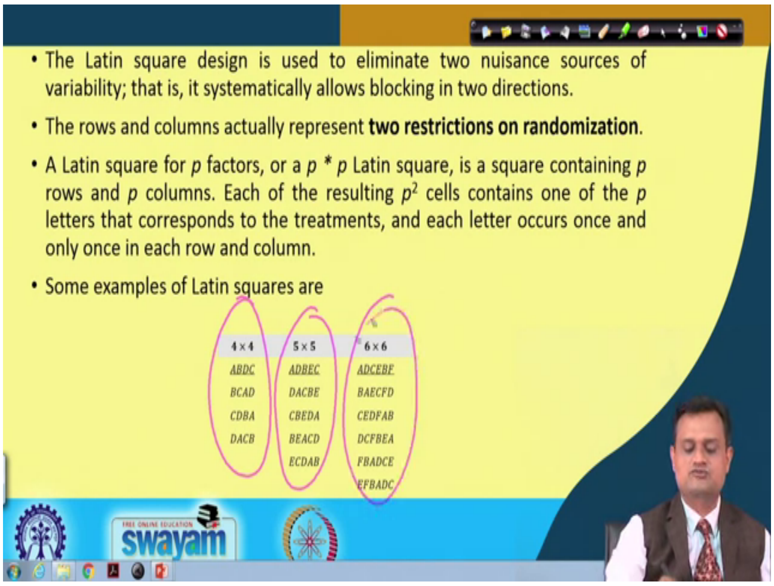
So, now just think that you have 2 nuisance factors to be controlled, now how you can really do it? So, when you have more than say 1 and 2 nuisance factor may be let us say each formulation you are preparing for a particular example, is in it is an outcome of some batch of raw material and it is also prepared by some operator. So, there could be the variability because of the operator and there could be because of batch you are dealing with 2 nuisance factors.

(Refer Slide Time: 25:56)



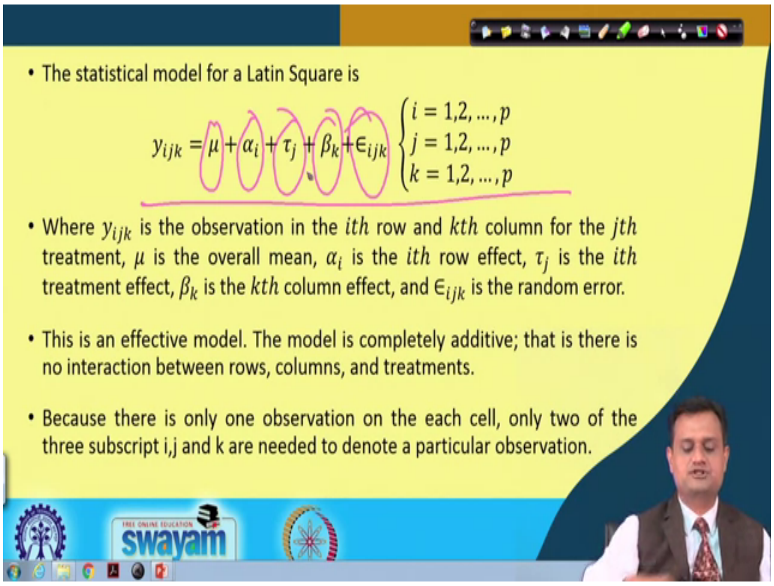
So, this is what I would say that I have 1 nuisance factor operator and I have another nuisance factor batches of raw material and what you can see that each particular operator can only make 5 say particular formulation and you have only adequate amount of say batch of raw material which can prepare only 5 say preparations or mixtures, now in this case I want to simultaneously analyze 2 nuisance factor; one is operator other is batches of raw material.

(Refer Slide Time: 26:38)



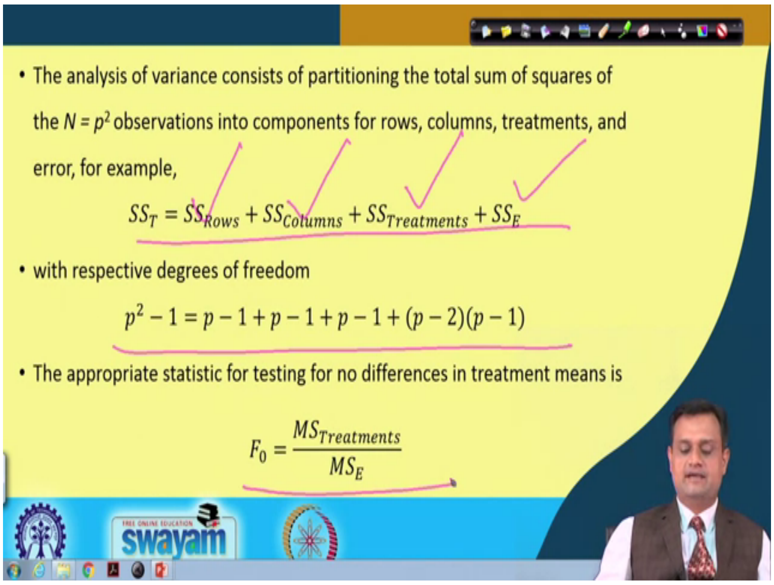
So, to do this you can use the Latin square design and you can have the Latin square design of 4 by 4 55 by 6 by 6 you can choose. So, we are not going into the different types of Latin square design, we are just trying to analyse.

(Refer Slide Time: 26:53)



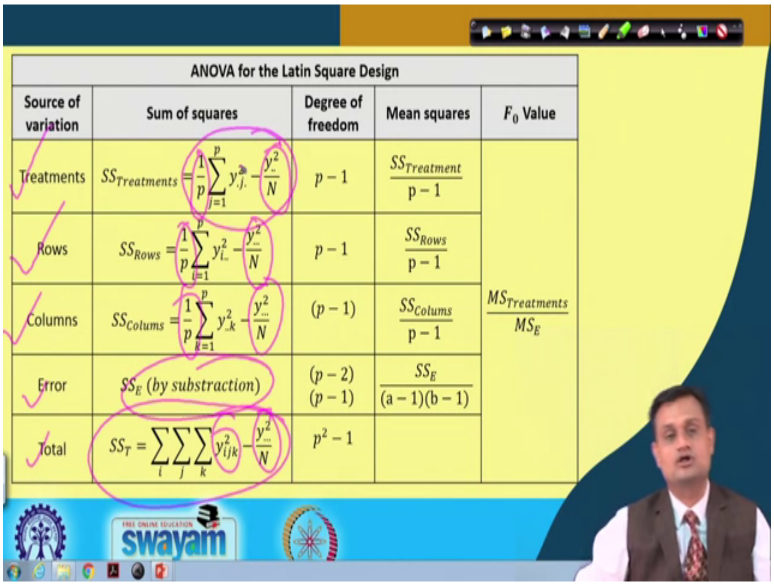
So, here my model would look like this yi j is basically the overall mean, then alpha i, that is the row effect tau j basically it is the treatment and you have the beta k that is the block effect and you have the epsilon ijk that is your error component.

(Refer Slide Time: 27:19)



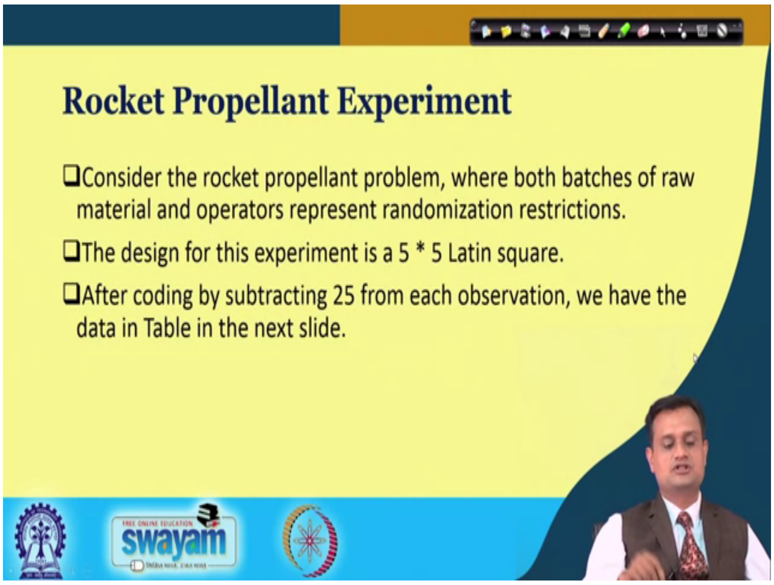
So, now with this we can have the equations like this SS T is SS row plus SS column plus SS treatment plus SS E. So, you have 1 nuisance factor another nuisance factor treatment and the SS error you can equate it this in order to get the degree of freedom and you have the statistics to be calculated MS treatment divided by MS E.

(Refer Slide Time: 27:47)



So, your particular ANOVA table will look like this, you have source of variation treatment row column and error 2 nuisance factor I am reminding you this expression will remain more or less same as we discussed here only you will see 1 by p 1 by p 1 by p and this you get by subtraction this is SS total, when you are subtracting this from grand mean each particular individual observation this component is common that is the grand mean square divided by total number of observation. So, you can easily interpret this and then I will calculate the F 0 value and 2 with the analysis.

(Refer Slide Time: 28:30)



So, this is the rocket propellant experiment.

(Refer Slide Time: 28:33)

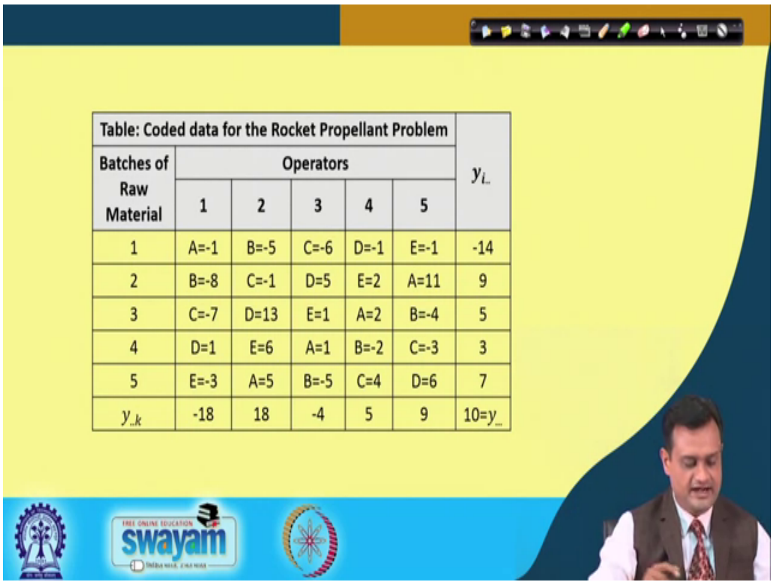
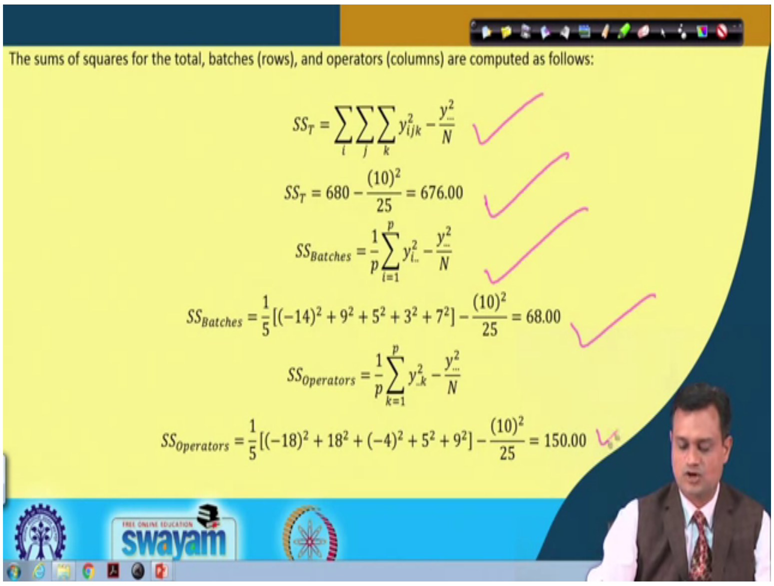


Table: Coded data for the Rocket Propellant Problem						
Batches of Raw Material	Operators					$y_{i..}$
	1	2	3	4	5	
1	A=-1	B=-5	C=-6	D=-1	E=-1	-14
2	B=-8	C=-1	D=5	E=2	A=11	9
3	C=-7	D=13	E=1	A=2	B=-4	5
4	D=1	E=6	A=1	B=-2	C=-3	3
5	E=-3	A=5	B=-5	C=4	D=6	7
$y_{..k}$	-18	18	-4	5	9	$10=y_{...}$

You have the data I am just trying to subtract each particular data which I have shown previously from the highest number data of that particular say column. So, let us say I did something like this, I have the highest reading 25 and that I am subtracting from the each observation.

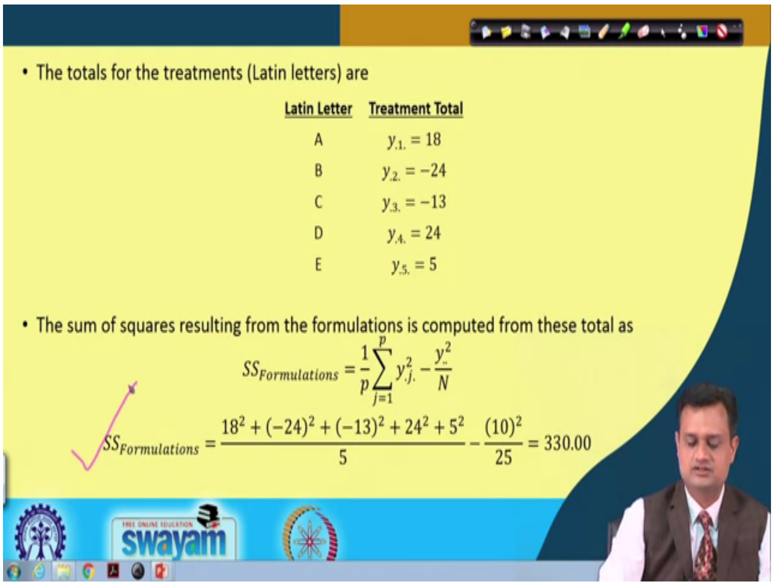
So, now what I get is the coded value and this is nothing, but by subtracting the highest value from each one. So, this is the same thing nothing else, but just for the sake of convenience easy computation I am doing this, I can have y i double dot I can have y double dot k, I have y triple dot and this values you can compute.

(Refer Slide Time: 29:26)



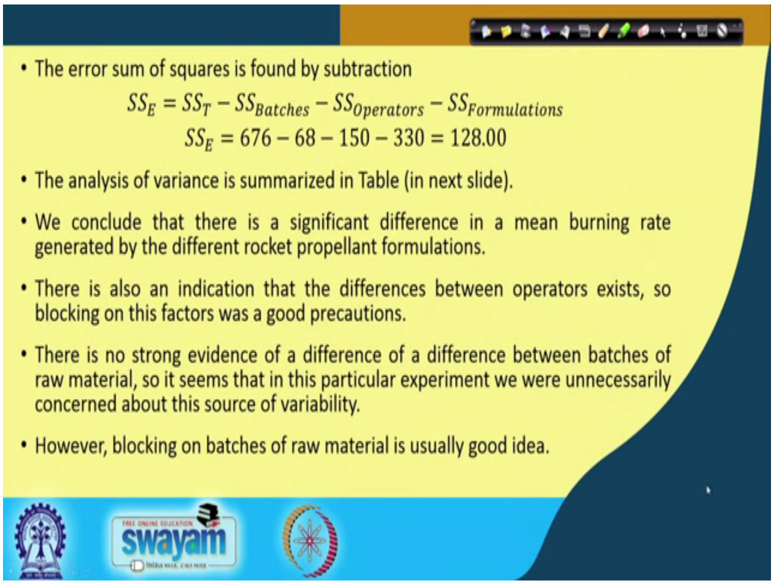
So, then you just plug in the values in the respective expressions and you will get SS T SS batches; SS batches and SS operator.

(Refer Slide Time: 29:40)



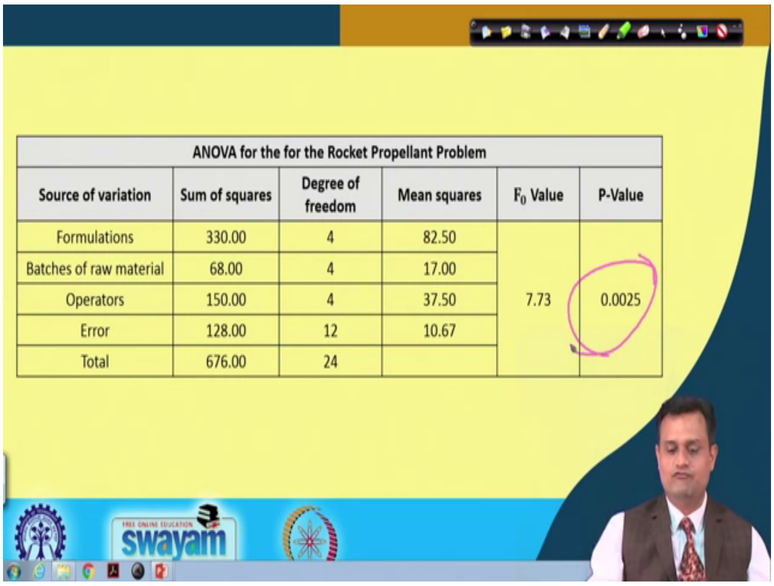
So, you have this values basically computed and then you also find the SS formulation. So, this is the expression for your SS formulation

(Refer Slide Time: 29:48)



Now, after doing this you can compute the SS E just by subtracting from the total and your ANOVA table looks like this.

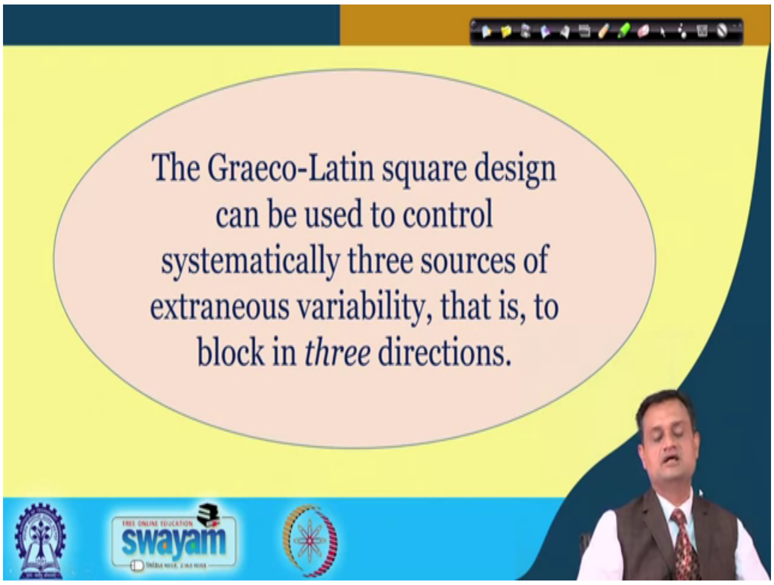
(Refer Slide Time: 29:54)



Source of variation	Sum of squares	Degree of freedom	Mean squares	F_0 Value	P-Value
Formulations	330.00	4	82.50	7.73	0.0025
Batches of raw material	68.00	4	17.00		
Operators	150.00	4	37.50		
Error	128.00	12	10.67		
Total	676.00	24			

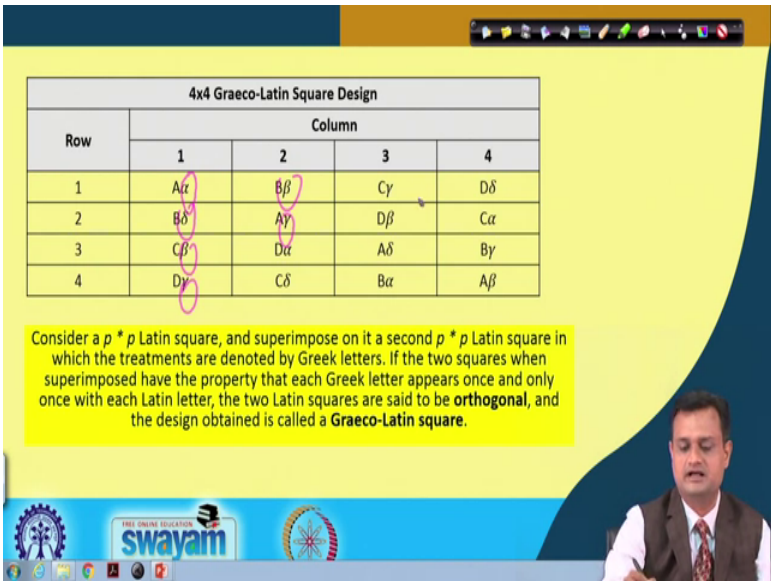
So, you see the p value and your P value is 0.0025, if you are operating with 0.05 as the level of significance, 0.0025 falls in the rejection region and you would say that my null hypothesis that my treatments are equal is rejected and yes there is an impact of the treatment formulation and this is where I check my model say results.

(Refer Slide Time: 30:36)



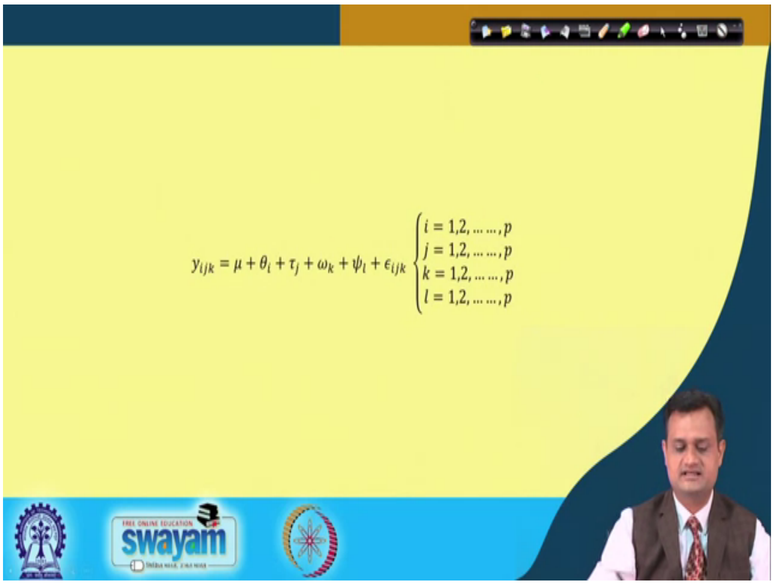
Now, there is another say approach improved version of the basic RCBD and this is called Graeco Latin square design. So, now think that you have 3 nuisance factors to control and this is where you extend the design.

(Refer Slide Time: 30:46)



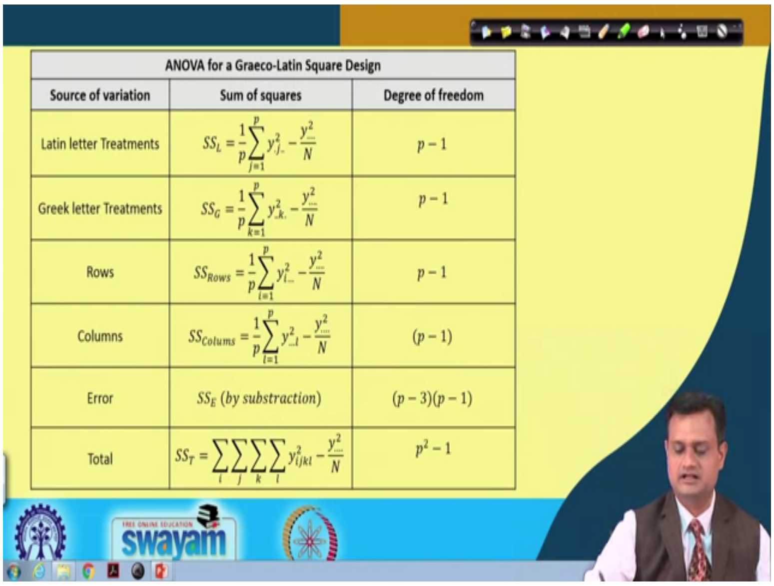
So, basically it is the p by p Latin square design superimposed on each other and you will see here the Greek letter which basically considers the third nuisance factor associated with each particular treatment.

(Refer Slide Time: 31:05)



So, the remaining procedure remains same my model looks like this, mu plus theta i plus tau j plus w k epsilon. So, you have 1 factor specific to treatment other is nuisance factor 1, nuisance factor 2, nuisance factor 3 and you have the error component.

(Refer Slide Time: 31:22)

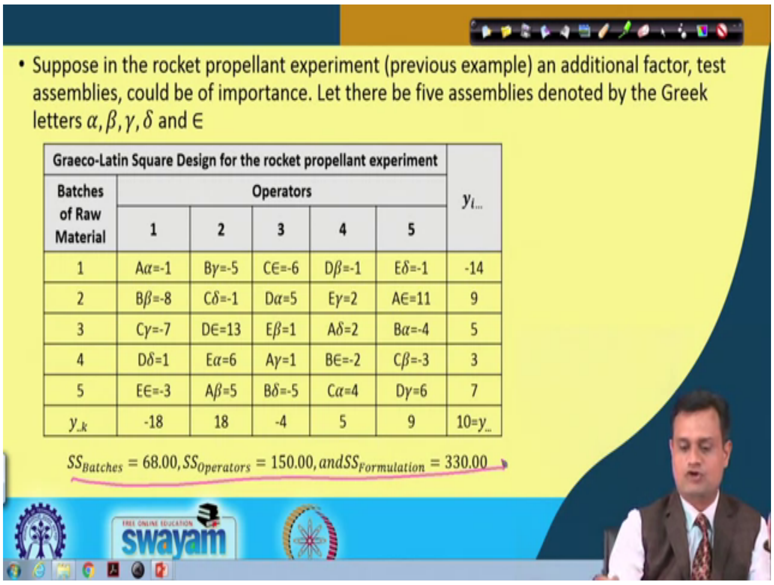


ANOVA for a Graeco-Latin Square Design

Source of variation	Sum of squares	Degree of freedom
Latin letter Treatments	$SS_L = \frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{....}^2}{N}$	$p-1$
Greek letter Treatments	$SS_G = \frac{1}{p}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}$	$p-1$
Rows	$SS_{Rows} = \frac{1}{p}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{N}$	$p-1$
Columns	$SS_{Colums} = \frac{1}{p}\sum_{l=1}^{p} y_{...l}^2 - \frac{y_{....}^2}{N}$	$(p-1)$
Error	SS_E (by substraction)	$(p-3)(p-1)$
Total	$SS_T = \sum_i\sum_j\sum_k\sum_l y_{ijkl}^2 - \frac{y_{....}^2}{N}$	p^2-1

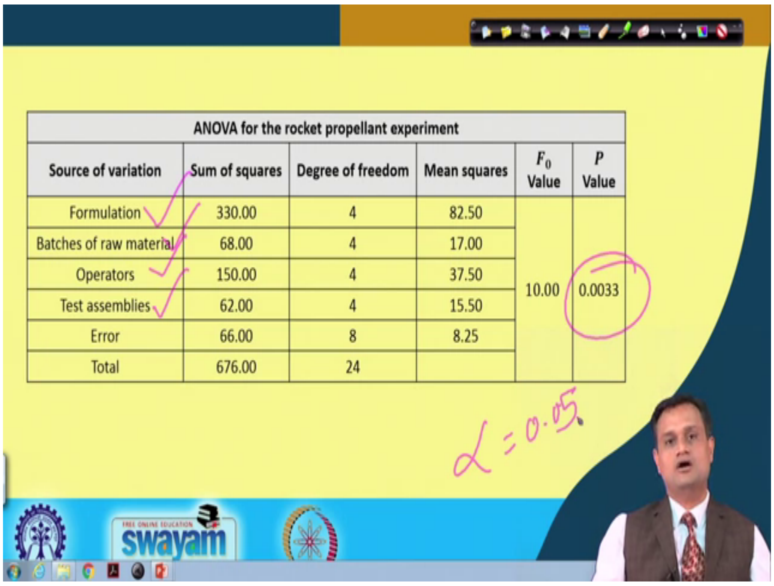
So, this is very simple to appreciate. Now, your ANOVA model will little bit get extended, Latin square treatments, Greek letter treatments, rows, columns errors and total these are the degrees of freedom.

(Refer Slide Time: 31:42)



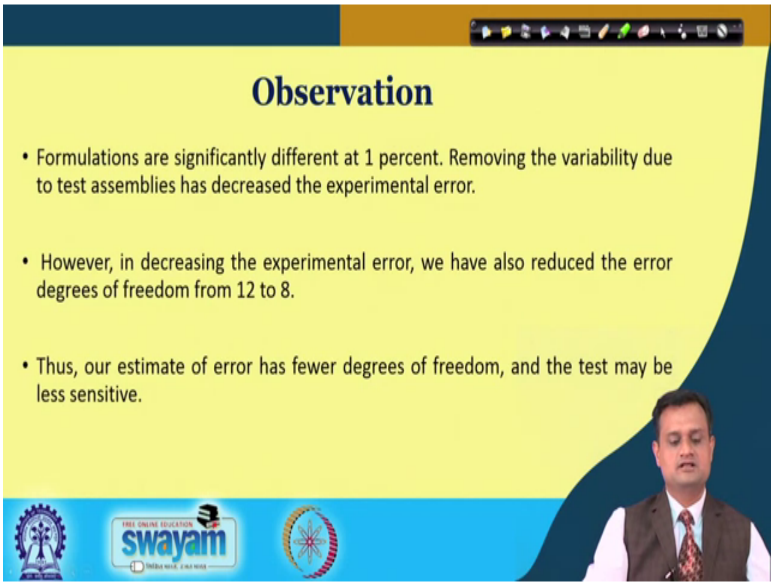
And with this expression you can once again solve the rocket propellant experiment, I am just trying to assign alpha beta gamma delta and then I am trying to find the values of SS batches, SS operator SS formulation.

(Refer Slide Time: 31:59)



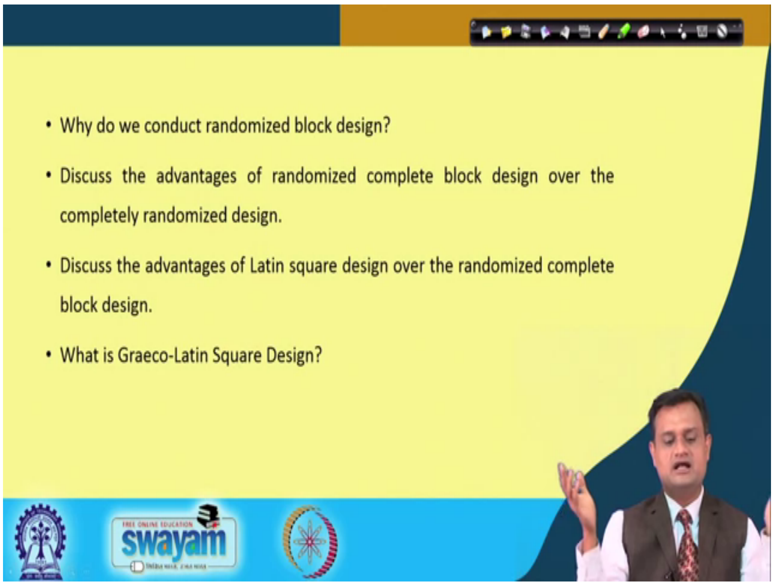
So, your final table after competition would look like this. So, you have basically formulation batches of raw material, operator test assemblies this is the additional factor nuisance factor I have added and what I get is the P value 0.0033. If you are operating at alpha is equal to 0.05, then this value is less than 0.05. So, it is in the rejection region I would say that yes, there is an effect of the formulation on the quality and my null hypothesis is basically rejected.

(Refer Slide Time: 32:37)



So, this is what we do. So, formulations are significantly different at 1 percent or 5 percent and this is what you can conclude.

(Refer Slide Time: 32:48)

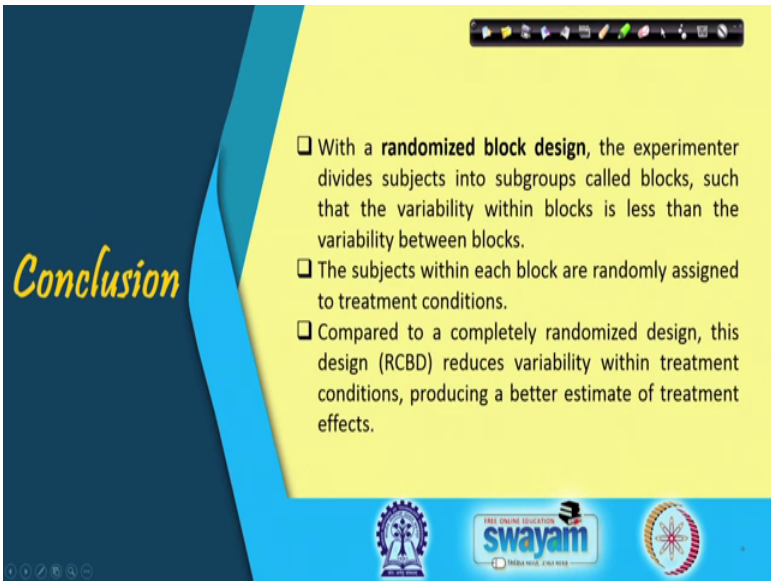


So, before I end let me plot couple of think it why do we conduct randomized block design when you have the randomized design ANOVA design available, discuss the advantages and disadvantages of this randomized complete block design. What is prompting us to go for the Latin square design and the Graeco Latin square design and typically what exactly we try to do compared to the randomized complete block design in Latin square design and the Graeco Latin square design.

(Refer Slide Time: 33:20)

These are the references mainly I am referring Montgomery.

(Refer Slide Time: 33:23)



So, RCBD basically helps to block the nuisance factor and hence you can have better analysis of the treatment and its impact on the response variable.

Thank you very much, keep revising apply for the real life application be with me enjoy.