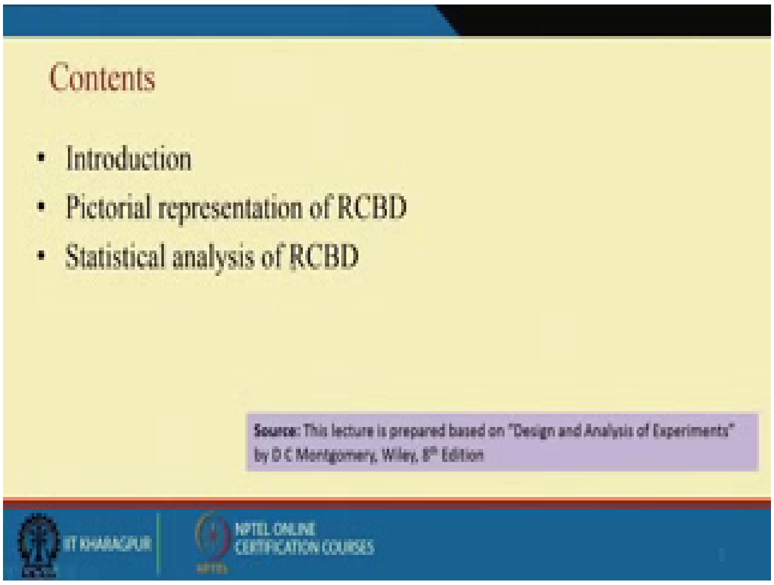
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture - 25
Randomized Complete Block Design (RCBD) : Introduction

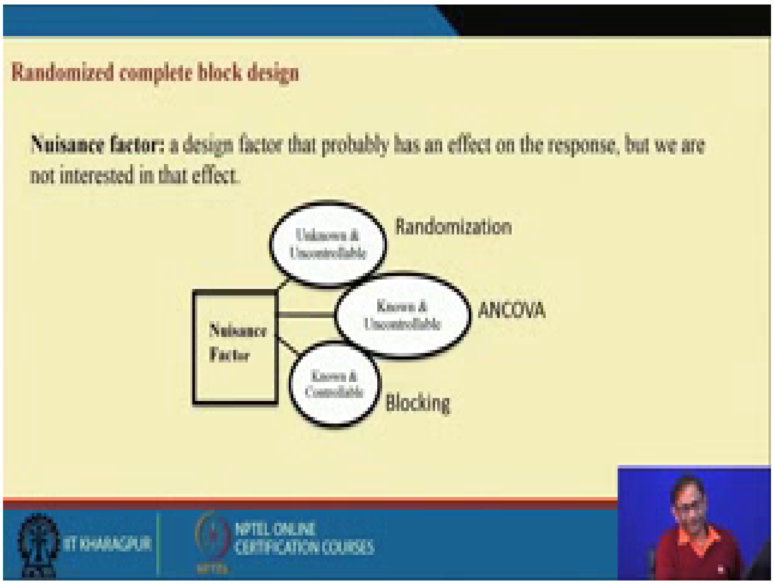
Welcome, we will start one new topic that is randomized complete block design. So, we will be spending around 2 and half hours on this topic and in this lecture we spend around 30 to 40 minutes of time to introduce randomized complete block design abbreviated as RCBD and the content of this lecture presentation R, we will first tell you when blocking is needed that is under introduction.

(Refer Slide Time: 00:52)



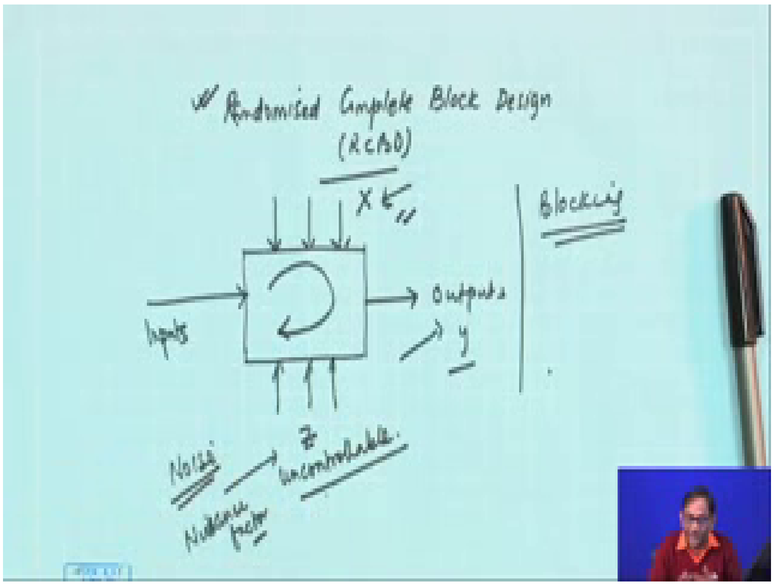
Then what is the pictorial representation R C B D, how R C B D can be represented in terms of table some layout and then one part of statistical analysis that mean how do you partition the data as well as the total variability into different component, like the way we have done in one way analysis of variance anova.

(Refer Slide Time: 01:38)



So, first understand important thing that the process model what we have discussed in earlier.

(Refer Slide Time: 01:52)



Will repeat this it take inputs and convert to value added outputs, suppose the output characteristic is y, this y also governed by set of controllable factors x and set of uncontrollable factors z and this conversion process a x inputs to value added output.

Now, these uncontrollable factors and sometimes we basically we can control, but we will control because of cost and other things, but whatever may be the whether controllable or uncontrollable in a situation sometime they actually are treated as noise. So, this is basically nuisance factor, nuisance factor.

Now, this nuisance factor cannot be ignored all the times sometimes because of randomization that their effect can be nullified, but many a times what happen there effect are there and if you do not do careful design for the experiment and accordingly. If you do not do experiment those noise or nuisance factors effect will be there and you will you will estimate the effect of the treatment with reference to controllable factors will be biased or other way other sense will be incorrect.

So, as a result it is essential in many situations to block the effect of nuisance factor and then conduct a experiment by blocking the nuisance factor and then get the data and analyse the data in with incorporating the nuisance factor into consideration so, that the treatment effects can be estimated while controlling the nuisance factors.

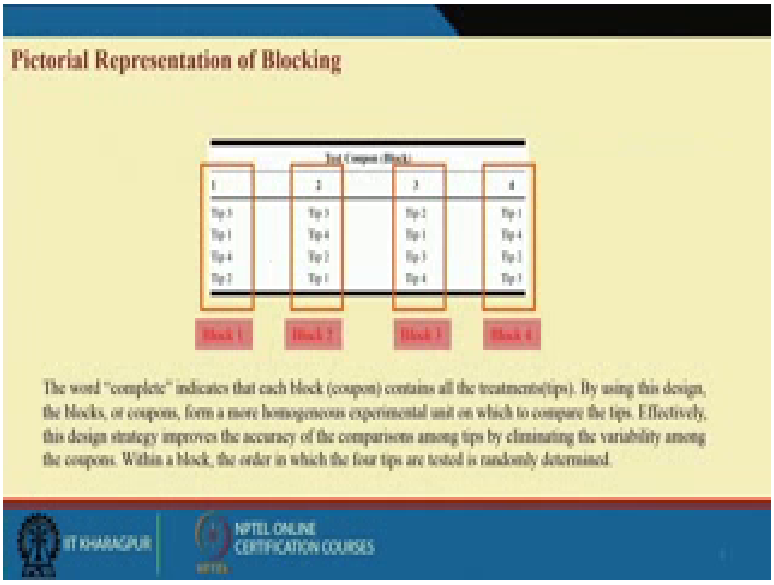
So, the important concept under such cases known as blocking, by blocking we mean that we will block the effect of nuisance factor during experimentation, we will you experiment in such a manner that the effect can be isolated ok.

So, now see the nuisance factor can be unknown and uncontrollable. So, you cannot do anything because you do not know what are those factors. So, only randomisation and even some time having more number of data that can that can help you in eliminating that effect, but many a time nuisance factors are known sometimes they are uncontrollable. So, when nuisance factor known and uncontrollable, the analysis technique known as analysis of covariance will help you in controlling the effect of uncontrollable nuisance factor and then estimating the effect of the control factors controllable factors.

But when sometimes it is known as well as controllable then it is better to block them, means control by blocking. So, today's introduction lecture will be on randomized complete block design, you know the blocks you know the controllable factors and you will block the nuisance factor in such a manner and also conduct the experiment in such a manner that under every block that you will be able to do experiment considering all the levels of the controllable factors.

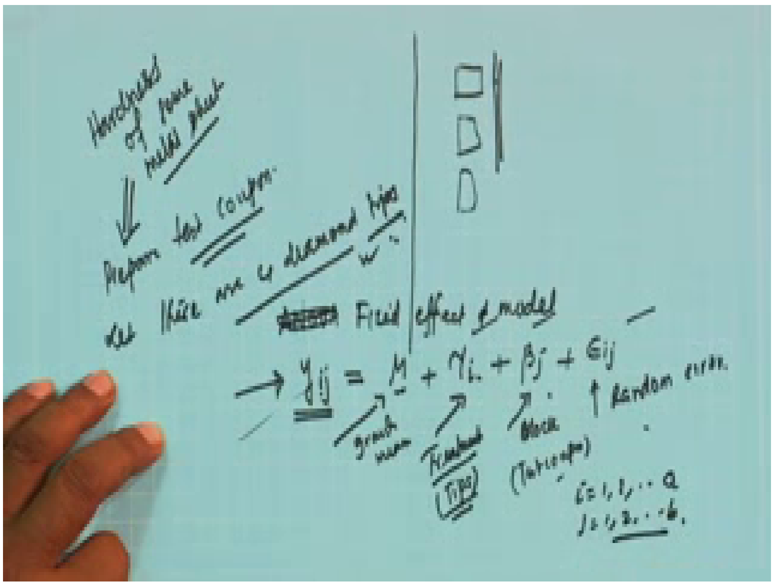
And you will be having sufficient number of experimental runs through proper randomization. So, that both the randomization blocking and replication these 3 principles will be adopted adequately and the data all together it will it will represent the actual condition under which the experiment should have should be done ok.

(Refer Slide Time: 07:02)



I will you one example here, before giving you the example I want to tell you that this lecture also repaired based on the materials taken from the book written by Montgomery design analysis of experiments.

(Refer Slide Time: 07:39)



So, here the experiment is that suppose you want to, you want to measure the hardness of some metal, metal sheet let it be and what happened.

So, you will have from this metal you can prepare test coupon, prepare test coupon every coupon is basically from a piece of this metal sheet. Let there are 4 diamond tips, let there are 3 diamond tips. So, this diamond tips basically this will be used to, you know that the hardness measurement. So, the diamond tips the penetration of the tip on the metal coupon if it is, if it is fall from certain height the penetration will give you a measure of the hardness. So, let this is the situation.

Now, is it may so happen that. So, you use different test coupons and suppose 5 test coupon for tip 1, 4 test coupon for tip 1 another 4 test coupon for tip 2 and tip 3, tip 4 like this and then if you analyse the data you get some results, but that may be erroneous because the test coupon which has be may be prepared of the metal sheet of different quality in terms of hardness.

So, under such situation what will happen may be the data will show you that the test coupon differs, but that difference may be due to the material effect or the test coupon effect rather than the tips effect. So, this kind under this situation what you do you block the test coupons, you create test coupon may be from homogeneous set of the sheet metal and then suppose 4 test coupon you prepare and from a homogeneous set that is a block 1 from second homogeneous set another 4 test coupon third homogeneous set another test coupon and then in each homogeneous set use all the tips and assign the tips randomly definitely.

So, then what will happen, you will have a pictorial representation of blocking like this. You see, this block one with reference to that metal the use hardness to be tested is homogeneous. Block 2 is another homogeneous set there may be variability or the difference in hardness between block 1 and block 2, but that is not of importance our important thing is that we want to know whether all the tips are giving you similar measurement or not so treatment here is the tips.

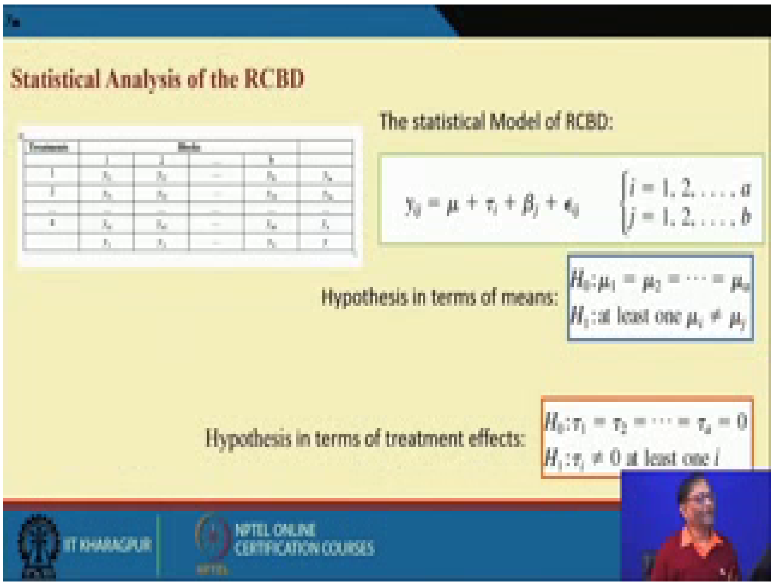
But if you ignore the coupon effects material effect you will get erroneous results. So, during design that is why you are creating homogeneous sets of sets of blocks and an each set of blocks will be used for all the treatments or will be attributed to or will be administered to all the treatments, here all the tips will be used.

So, material that representation is like this, block 1 see all the 4 tips were used, block 2 4 tips were used, block 3 4 tips. block 4 4 tips were used. So, this is what is known as that randomized complete block design, the word complete indicates that each block contents all the treatments. You see in every block, your block with respect to test coupon in every block all the 4 tips were used by using this design the blocks or coupons from a more homogeneous experimental unit on which to compare the tips.

Effectively the design strategy improves the accuracy of comparison among the tips by eliminating the variability among the coupons. Within a block the order in which the four tips are tested is randomly determined; that means, blocking within block randomization and also sometimes what happen you may go for replications also more than one, more than one runs experimental runs ok.

So, under this as we are not interested in knowing that how the materials are effecting rather we are interested in knowing that whether tips are effecting in the measurement. So, may be under each block one replicate is sufficient for treatment because effectively the number of blocks will be used as some kind of replication for each of the treatment levels.

(Refer Slide Time: 13:33)



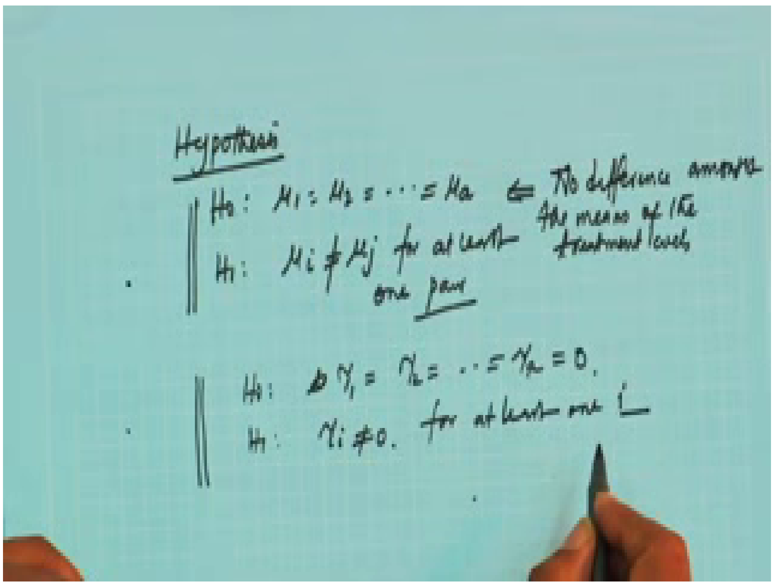
Then what will be the data? Data will be like this in general you your treatment will be having a number of levels suppose there are b number of blocks and suppose for each block treatment combination you are doing one experimental your conducting one experiment only. then you will be getting data y 112, y 1 b something like this and this side this is that y 1 naught is the total rho total y 2 dot is the second rho total like this y a dot 8th rho total and in the similarly in the column also column totals corresponding column totals y dot 1 first column total, y dot 2 second column total, y dot b 3rd column bth column total and y double dot this is basically the grand total so fine.

But essentially what you are getting, you are getting data y 1, y 2 this a treatment 1 2 a blocks 1 2 b this data set you are getting. So, then what is the model here your any observation general observation which can be written at y i j.

So, any general observation y i j this can be written in partition or written like this it has the grand mean effect plus tau y that is the treatment effect, plus beta j this is the block effect, plus epsilon i j this is what is the random effect random error. So, with reference to the example that mean treatment it is the tips whether the tips give effect is there in the measurement blocks are nothing, but the test coupon and random.

So, a general observation having grand average or grand mean component, treatment effect, block effect and random component, this is what is the anova model or general what I can say that fixed effect model, fixed effect data effect model. Why we are considering the treatment has fixed, any observation will be and these mu tau i beta j these are all the parameters of this model, here i stand from 1, 2 a and j stands from 1, 2 b ok.

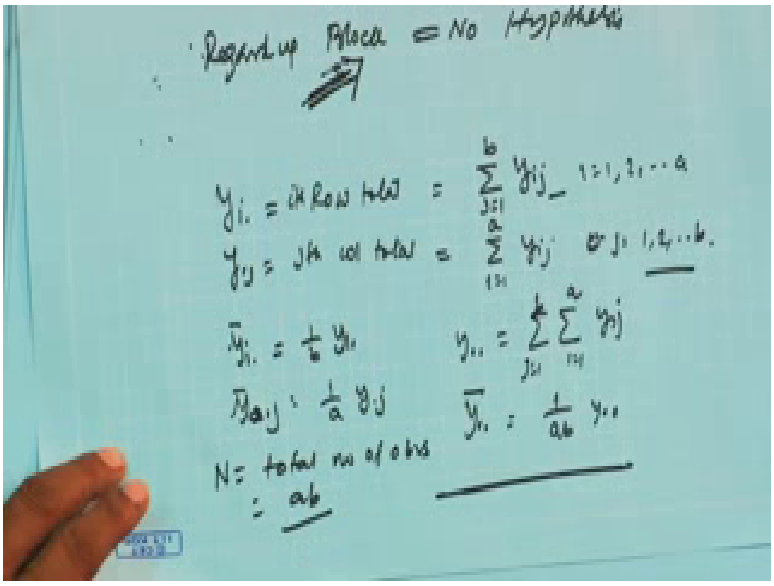
(Refer Slide Time: 16:45)



So, our hypothesis of interest is here h 0 that mu 1 equal to mu 2 dot dot dot equal to mu a. So, that mean there is no the treatment that mean no difference, no difference amongst the means. The names of the treatment levels and what is our alternate hypothesis mu i not equal to suppose mu j for at least 1 par. So, this is known as hypotheses for means, suppose if you want to use the hypothesis for the treatment effects then you write correspondingly tau 1 equal to tau 2 dot dot dot equal to tau equal to 0 and alternate one is tau i not equal to 0 for at least 1 i.

But please remember we are not making any hypothesis regarding block.

(Refer Slide Time: 18:16)

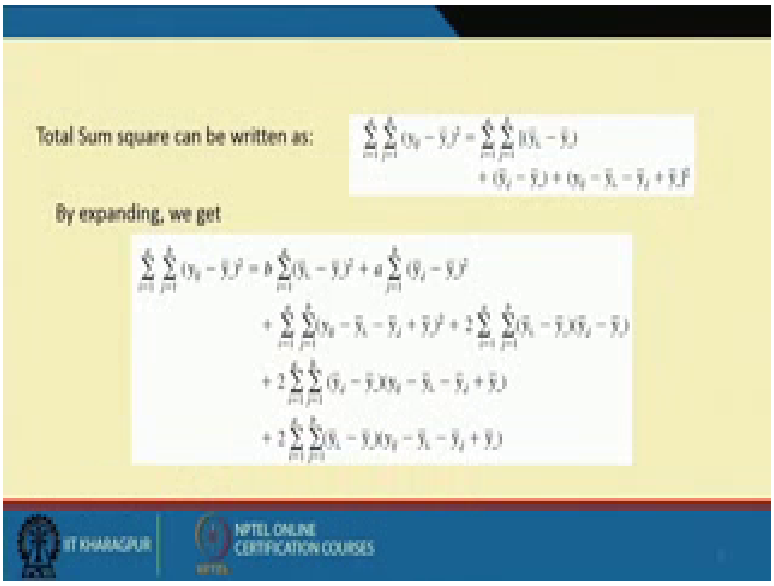


Regarding block no hypothesis, why? Because block is a block is here the test coupons are known and there is variability they there is variability and we are not interested to know how the different test coupons having different kinds of variability whether we are interested in the treatment effects that is the tips effect. So, for block or here in with this example the test coupon effect we are not making any hypothesis we are not interested to do this.

So, under such situation the calculations will be like this, what is in general, why i dot means that is rho total ith rho total, then y dot j in jth column total and that you know. So, when you are considering rho total that mean you are going across all columns. So, j equal to 1 to b that y i j and i equal to 1, 2 a. So, when you are going for across for a column you are going across i equal to 1 to a y i j, i j equal to 1, 2, b

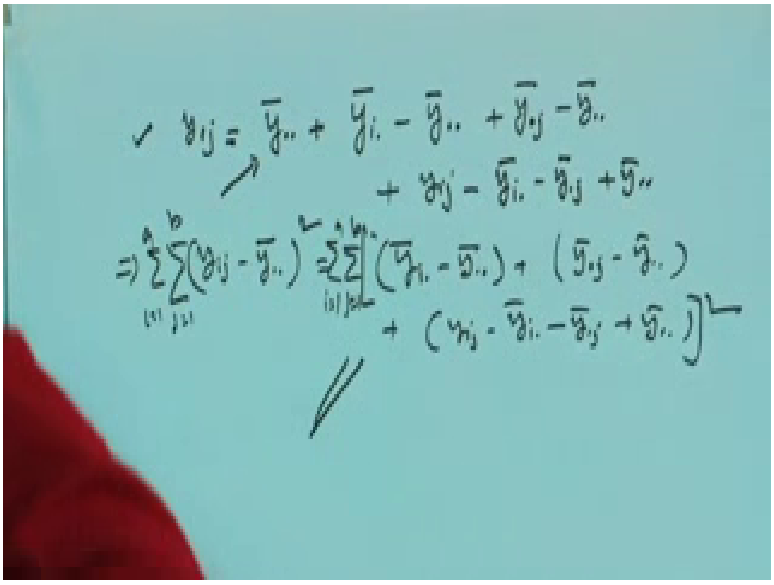
Then if I want the average y 1 dot average. So, it will be 1 by b, y 1 dot, if I want y dot j average this will be one by a y dot j, suppose I want the total observation total, grand total then i equal to 1, 2 a j equal to 1 to be y i j and if you want grand total this will be 1 by a b y dot dot. So, these are the way from the calculation part and ultimately what is the total number of observation n equal to total number of observation equal to a into b ok.

(Refer Slide Time: 20:51)



Let us see that the individual when you collect data you have y i j equal to y double dot bar which is grand average plus y i dot bar.

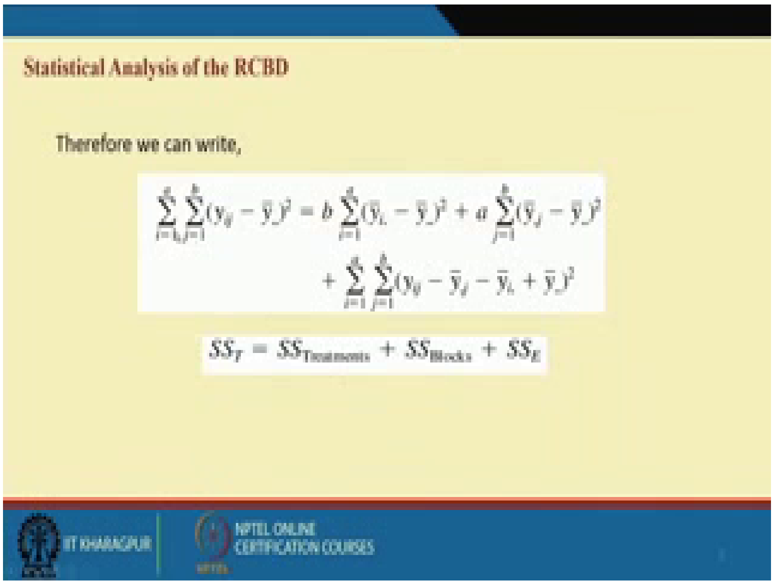
(Refer Slide Time: 21:01)



Suppose minus grand average plus y j dot bar minus grand average plus if i write y i j then if in order to make left hand and right hand equal say this and this will be cancel out. So, that mean minus y i dot bar minus y j dot bar plus y double dot bar.

So, we now if we do little modification y i j minus y dot dot bar then it will be y i dot bar minus y dot dot bar plus y dot j bar minus y dot dot bar plus y i j minus y dot bar minus y j dot bar plus y dot dot bar. So, you take some square, square each take j equal to one to b, take j equal to 1 to b take summation i equal to 1 to a, summation i equal to 1 to a and do algebraic manipulation and finally, what happen you see the slides you will see that these ultimately become first quantity b of these second quantity a of these third quantity these and there are some covariance type of things the 2 times of this. And if you take the sum you will find out that the 2 time this covariance kind of thing that will become 0 and as a result.

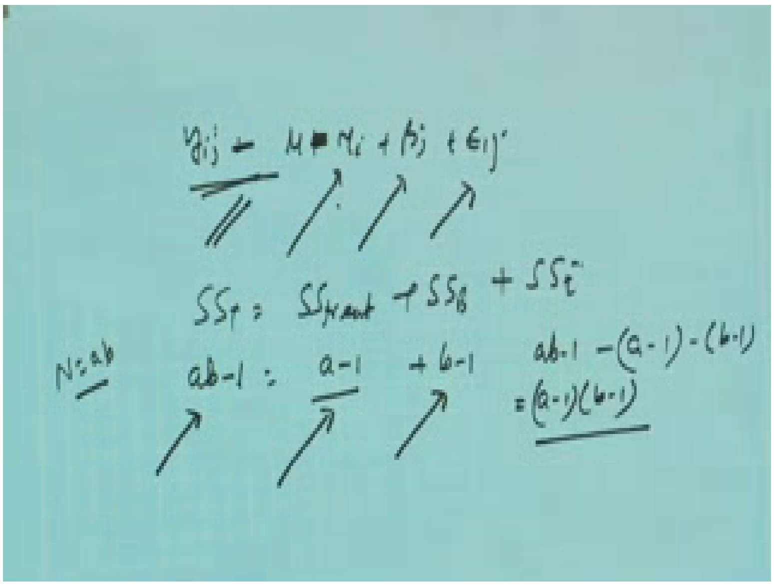
(Refer Slide Time: 23:00)



The resultant will be this is the sum totals sum, sum square total y i j dot minus this square equal to this plus this square plus this.

So, sum square total is now divided into sum square treatment, sum square blocks and sum square error that is what you have seen because your original data individual i y i j you have written mu plus tau i plus beta j plus epsilon i j.

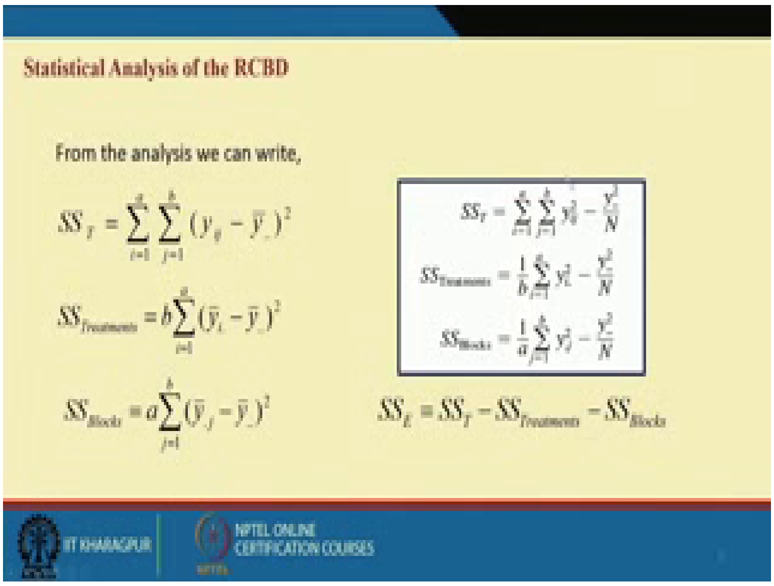
(Refer Slide Time: 23:25)



So, if I write this minus this equal to this then ultimately the total derivation is coming from the treatment point of view, block point of view, error point of view as a result SST equal to SS treatment plus SS block plus SS error and you will find out the degrees of freedom will be a b minus 1 because your n equal to a b. And here a minus one treatment plus b minus 1 and then rest will be a b minus 1, minus a minus 1, minus b minus 1 sum will be like this ok.

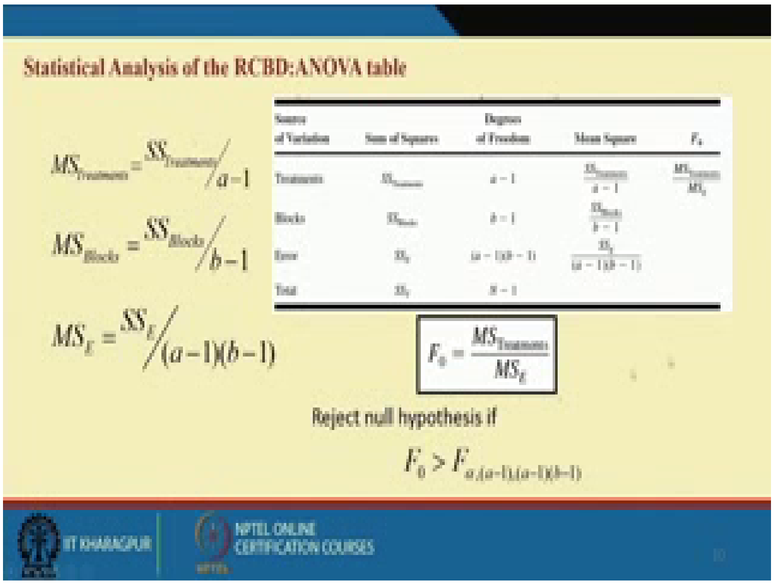
So, what happen you will find out that your error will be a minus 1 into b minus 1. So, degrees of freedom for total is this, treatment this, block this and this will become a minus 1 equal to a minus 1 into b minus 1 fine. Now, you require to compute SS treatment SS block SS c and that already you have seen this is SS treatment this is SS block, SS total SS treatment, SS block, SS error this can be written in this format that e g.

(Refer Slide Time: 24:57)



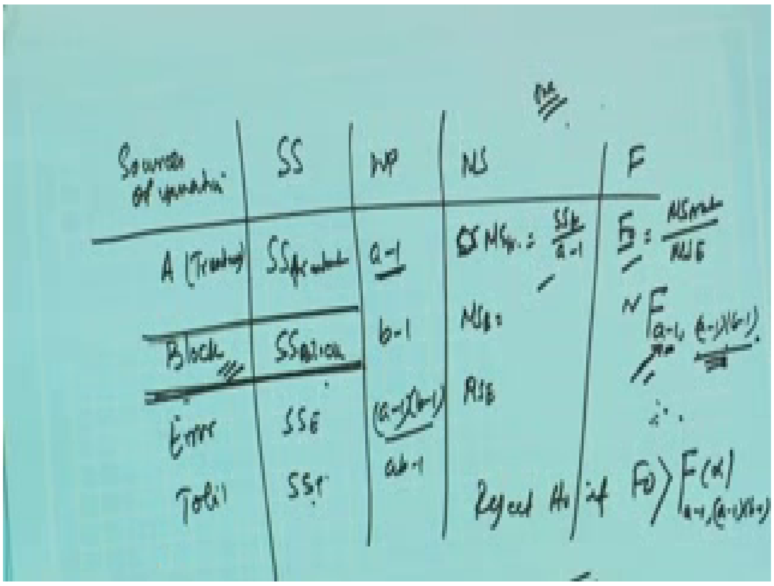
So, this can be this can be computed using SS t equal to this SS treatment this and. So, here you require average here you do not require average only the from the total square point view you are getting this kind of computation you have seen in one way one way anova also ok.

(Refer Slide Time: 25:20)



And then you what you require, you require to prepare the anova table please keep in mind in anova table first and first thing you have to do is sources of variation.

(Refer Slide Time: 25:29)



In R C B D with one factor sources of variation first one is a that is the factor, then block. So, if I say this is block, this is what is your treatment then error, then total you required to compute SS. So, it will be SS a or SS treatment, SS block, SS error, SS total you require d o f a minus 1, b minus 1, a minus 1 into b minus 1 a b minus 1.

Find out m SS s that is m s treatment equal to SS treatment by a minus 1 similarly m s block similarly m s error. Then find out f we are interested only in this first 1. So, f 0 is m s treatment by m s error which follows f distribution with a minus 1 degree of numerator and a minus 1, b minus 1 denominator degree of freedom.

See the difference between these and 1 factor complete randomized design, you will find out the one of the difference is that here a minus 1 is there earlier denominator degree of freedom is change they are a into n minus 1 and here a minus 1, b minus 1 and essentially the major difference is that in without blocking if you do this component you avoid eliminate you will not consider.

But when you are actually blocking it you are partitioning the sources of variation into box also. So, this will be taken out and as a result what happen the F 0 and other statistics value will become different and you will be getting the correct kind of result and then you will visit F 0, if F 0 this is greater than a threshold value of a minus 1, b minus 1 threshold value of alpha level of significance.

So, this is what is for today. So, I told you that what is R C B D. So, R C B D is randomized complete block design randomize what is there every experiment the water and the way material and the operator other things are assigned they must be random. Here block is coming because you know there are certain machine factors and those machine factors are known and controllable. So, you can block them during experiment we have given one example where tips versus the test coupon test coupon is block and different tips are treatments.

And then what we have discussed that we know that there are force there are different sources of variations with the example given with one factor the factor or treatment is one variation, block one block another variation, there will be random error that is another variation. So, you know the sources of variation and accordingly you are calculating s a s, m a s different degrees of freedom and F statistics here you are not interested to know whether the different blocks are different or not. So, that is why you are not conducting any hypothesis resting related to blocks, but you can do that also, but it is not required here. We will in the next class we will see one example on blocking and then there are many more critical issues in blocking those will be discussed.

Thank you very much.