Total Quality Management - II
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 22
Factorial Designs – VIII

Welcome back my dear friends, a very good morning, good afternoon, good evening to all of you. And this is the NPTEL MOOC course series; and the course is TQM II. We are in the fifth week; that means, you have already completed half of the whole number of courses, and this is the 22nd lecture. And as you know I am Raghunandan Sengupta from the from IME department IIT Kanpur.

So, to just to recap what we have done if you remember that for in each of the classes, I try to spend about 2 to 3 minutes trying to recap what we have covered in the last two or three lectures; and obviously, if required I go back to the initial lectures give the bullet points. So, if you remember what we have we were our main concern is there are factors, they can be a factor a, factor b, factor c, factor d whatever the numbers be. And we want to find out the interrelationship with the factors and the effect of the factors on trying to basically test some hypothesis.

So, whatever the hypothesis it can be say for example we think the mean values are equal for each factors or it may be say for example, one of the factors effect is dispersion or standard deviation is more than for the other. Or say for example, standard deviations for the dispersions for all the factors are equal and so on and so forth. Now, when you are considering the factors a, we basically try to find out the individual effects for a, for b, for c, for d, and then we try to find out the combined effects taking two at a time.

So, if there are say for example, capital N number of factors, it will be N c 2. And if they are same now if you consider the effect of the factors taking three at a time, it will be capital N c 3 and so on and so forth. Now, when we consider the effect of the factors is basically is to find out is the total errors coming out from the effect of the factors taken one at a time or two at a time or three at a time. Once we do that we try to find out what is the degrees of freedom. Once the degrees of freedom I found out, we find tried to find out what is the mean square. Once the d mean square is found out we find out the whether the half this is being rejected or accepted.

Now, it also should be remembered that we intuitively consider the distribution per say to be normal that is an assumption which may not be true, but we consider that in order to make our life simple. Then we also consider different type of other tests corresponding to through the fact that you want the mean values to be 0 for the factors then we found out the that so called test. So, all these are hypothesis considering the fact that the mean values of one of the factors whether corresponding to the dispersions of the factors which is tau suffix i beta suffix j gamma suffix k where i is equal to 1 to small a, j is got 1 to small b, k is equal to 1 to small c and so on and so forth.

So, those can be equal that not individually equal that means, tau is not equal to beta, but which tau i's are all equal for i is going to 1 to a, similarly for beta, similarly for gamma. And we consider different types of hypothesis. Later on we saw that say for example, if the we and based on that later on we saw that how the graphs could be drawn, how the factor planes could be drawn in order to find out whether there was linear relationship or non-linear relationship. Later on yesterday for the last class; yesterday of what I am using for the last class which is this 22nd one we did also consider that if the relationship was between a quantitative factor and the qualitative factor.

So, let me come to the example of say for example, the battery. So, what you have is the or say for example, the etching problem. So, when you consider the battery example there were some temperatures which could be quantified, and there were materials which could not be quantified. So, they were basically a combination of qualitative and quantitative. It was all quantitative then trying to basically do a so called regression modeling considering the effects or the relationships are given by a simple regression modeling was easy because all the quantitative facts could be converted into a simple regression model and we can convert the problem accordingly and solve it accordingly.

Now, in case if there are quantitative our main task is to find out that how we can basically do a mapping, this mapping I am not talking from the method with a mathematical point of view. It is basically a more very simple from a Layman's point of view considering say for example, there are three materials we will consider those three materials can be quantified in the scale of say for example, minus 1, 0 or plus 1, or it can be minus 2, minus 1, 0, plus 1 and plus 2 depending on whatever number of number of factors they are. That means, under each factor if the variations is very large that means, that there are many different type of mm factor levels let me consider use the word levels.

And then we saw that that considering that if you have been able to convert the qualitative one the quantitative one; you can replace again very simply in the models we have discussed in a simplistic sense. And try to find out what is the values of f naught, f naught means depending on what is the hypothesis, and then we can whether we can accept it or reject it. So, let us continue discussing that further in the 22nd lecture.

(Refer Slide Time: 06:09)

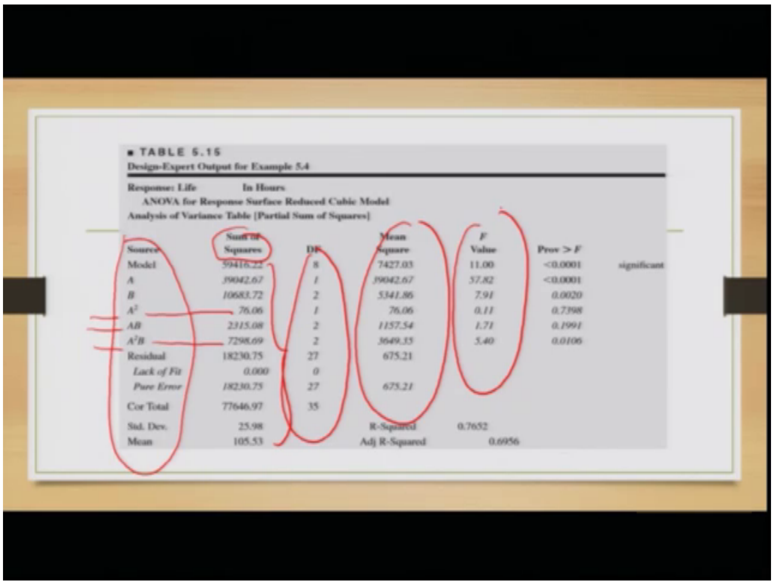


TABLE 5.15
Design-Expert Output for Example 5.4

Response: Life In Hours
ANOVA for Response Surface Reduced Cubic Model
Analysis of Variance Table [Partial Sum of Squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prov > F	
Model	59416.22	8	7427.03	11.00	<0.0001	significant
A	39042.67	1	39042.67	57.82	<0.0001	
B	10683.72	2	5341.86	7.91	0.0020	
A^2	76.06	1	76.06	0.11	0.7398	
AB	2315.08	2	1157.54	1.71	0.1991	
A^2B	7298.69	2	3649.35	5.40	0.0106	
Residual	18230.75	27	675.21			
Lack of Fit	0.000	0				
Pure Error	18230.75	27	675.21			
Cor Total	77646.97	35				

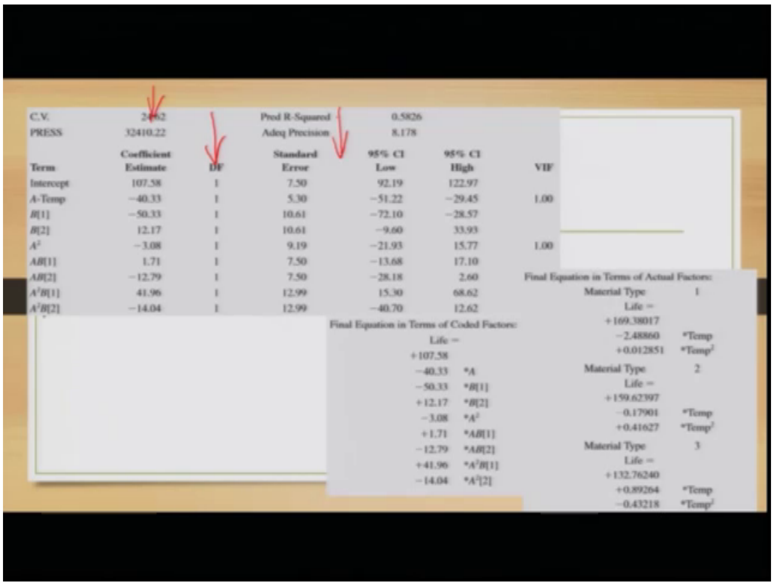
Std. Dev.	25.98	R-Squared	0.7652
Mean	105.53	Adj R-Squared	0.6956

So, if you consider the design m expert output for the example 5.4 so which is the ANOVA model. So, the as we remember and I have been all always discussing the first column would be all the factors, so it will be A, B, C if there are three. Then it would be AB, AC, BC, then ABC then there is in the total one.

The second column would basically be all the total sum of the squares corresponding to this factors; that the third column is the degrees of freedom corresponding to this factors. then you have the mean squared error whereas, corresponding this. And the last value basically be the f values, and the p values. P means values means the degrees of freedom and level of confidence which you have based on which you try to basically either accept or reject the null hypothesis.

So, in this you have the source of the models and then you have AB. Now, here if you remember I did mention that that some of the problem formulations would be such that rather than finding or the linear relationship it could be quadratic relationship also.

(Refer Slide Time: 07:13)



So, in this case what you have sorry for that in this relationship what you are doing if you pay attention to the first column - this one. So, they are factors so called factors which will consider which are basically now not AB, apart from AB, you have basically a square you can have A square B, you can have A B square. Based on this you can find out what is the effect; so the sum of the squares, which is the effect of the errors for all these terms are given. So, basically for A square the sum of the square is 76; for A square I would not repeat the common one which we have been discussing not AB. We will discuss A square B the sum of the squares is 7298. So, I am leaving aside the decimal.

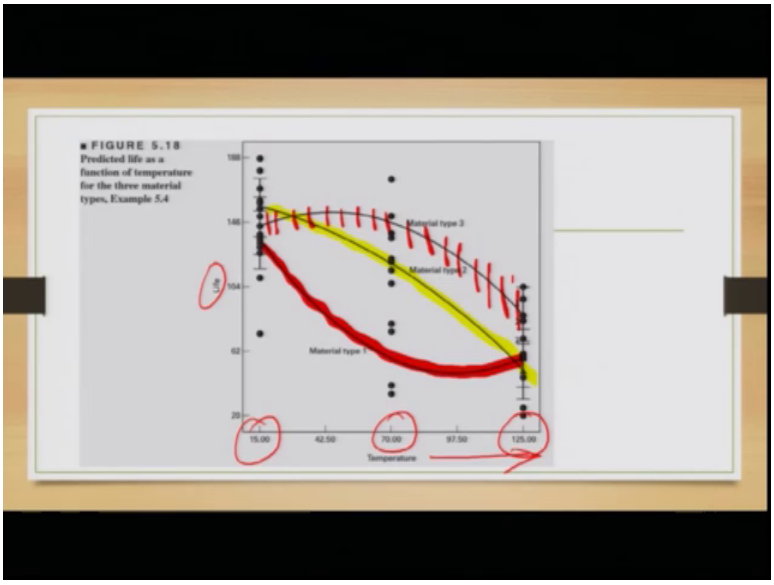
Similarly, you would also have some sum of the squares corresponding to AB square and so on and so forth. Then you have the degrees of freedom, once the degrees of freedom you can find out the mean squares, one the mean square is how well you can find an F values and basically come to the conclusion whether you want to accept or reject it.

Now, corresponding here when you are putting the say for example, in regression model very simply, you would basically have the R square and the adjacent R square. So, adjacent R square would basically consider degrees of freedom, and you can find out the values and basically come at intelligently about the problem.

So, when you do the problem related to the temperatures example, so you basically have temperature then you have basically reading it along the first column and only concentrate on the important factors it is A square then you have A square B and so on and so forth. So, you have the question coefficient estimate the degrees of freedom the standard errors.

I am just giving you the highlighted points you have the coefficients estimates, you have the degrees of freedom, you have the pre adjusted R square. And based on that once you find out the coded factors the decoded factors would be given by the values whether they are plus or minus and then you basically solve the problem accordingly.

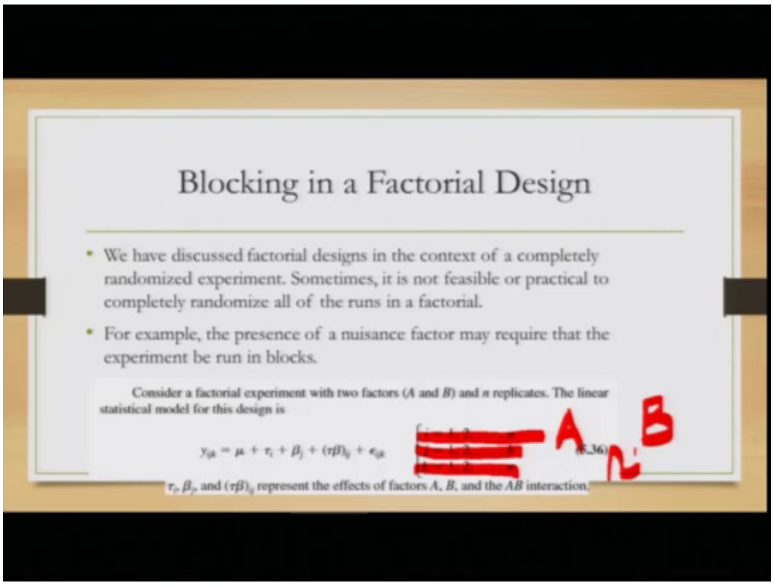
(Refer Slide Time: 09:23)



When you try to find out the predicted life so with respect to the temperature, so there would be different materials, materials would basically corresponding to different factors which you have. So, for the different materials, you will have the temperatures if you remember there were 15, 70 and 125. So, they are plotted along the x-axis. So, this is 15, this is 70, this is 125. And based on that if you have the life, life was basically in hours working, so you have the life.

So, different materials will give you different relationships where the linear or non-linear. So, in this case, I will use a different color to highlight that. So, this part is almost linear. So, if you consider the other part, this one is non-linear, but it basically decreases as temperature increases. And the other part, I will try to use a different light color if possible it is not coming I use hash. So, this one basically is increasing till one 70 degrees and then basically start decreasing from 70 to 125.

(Refer Slide Time: 10:36)



So, now we will consider the concept of blocking in factorial design problem. So, you discussed the factorial design in the context of completely randomized experiment. So, it was a random experiment factors were decided that data set was considered considering a random experiment, and we did R, R analysis to either to prove or disprove the null hypothesis accordingly.

Sometimes, it may not be feasible or it is not feasible or practical to completely randomized all the runs in a factorial model, because it may not be practical or it may not give you the exact answer. So, for example, the presence of nuisance factors which would be there may require that the experiment be run in blocks.

So, same consider that during the you are doing an experiment. And for temperature 125 at some particular time the humidity is very high which is not a factor to be considered which means the external effects white noise is a nuisance parameter value the effects are very high. So, you want to basically do away with it. Or say for example, a very low temperature the humidity is very low or the pressure it suddenly becomes very low, and the environmental pressure based on which you are doing the experiment.

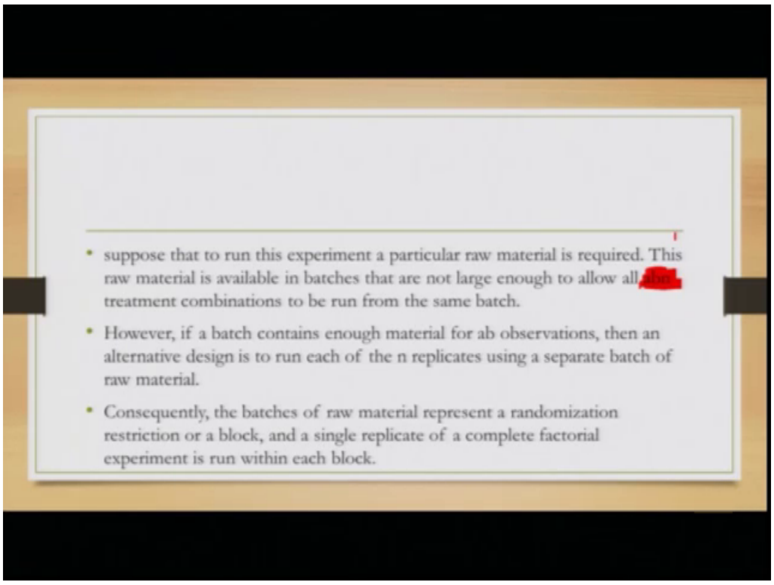
So, in that case, you would basically need to have a control on the experiment in order to do away with the external effects. So, consider so let us consider a a factorial example with two factors A and B. So, it would be AB and the combination of A B would also come A and B separately in the combination of a becoming. The linear statistical model would be giving like this which is again we I will repeat because you have done it, I am sure you have understood very well, but I will repeat it in order to make it again very specific to the problem the general concept remains the same.

So, what you have is basically the values, the values are given by y. So, the if they are a say basically two factors A and B, the subscripts would be i for a, j for b, and k would be for the values from k is equal to 1 to small n, so that is why you see y suffix ijk. Then you will basically have the mu, mu is basically the average of the average of the average. considering that you are plane trying to find out the average for all n s for all the a s and for all the b s. So, there will be three averages; obviously, you replace that mu value with mu hat and try to do your example accordingly, because mu hat is basically from the sample.

Then you have basically tau y corresponding to the so called dispersion with respect to factor A. Then you have basically beta g corresponding to the dispersion, dispersion of the movements from the average of the average of the average values corresponding to factor B. Then you have basically tau b suffix i j corresponding to the fx or both of them. And finally, if the error term which is epsilon ijk depending on whatever values of i's are there j's are there which in reading you are going to take.

So, as mentioned here and as you can note down I will just highlight it; i is equal to 1 to small a for factor A; j is equal to 1 to small b for factor B, and k is equal to 1 to small n for this number of observation which is n. So, here tau y beta j and tau beta ij represents the effects of the factors A, B, and AB intersection taken separately. So, they are for tau is basically for A, beta A is basically for B, and tau beta is basically for AB.

(Refer Slide Time: 14:09)

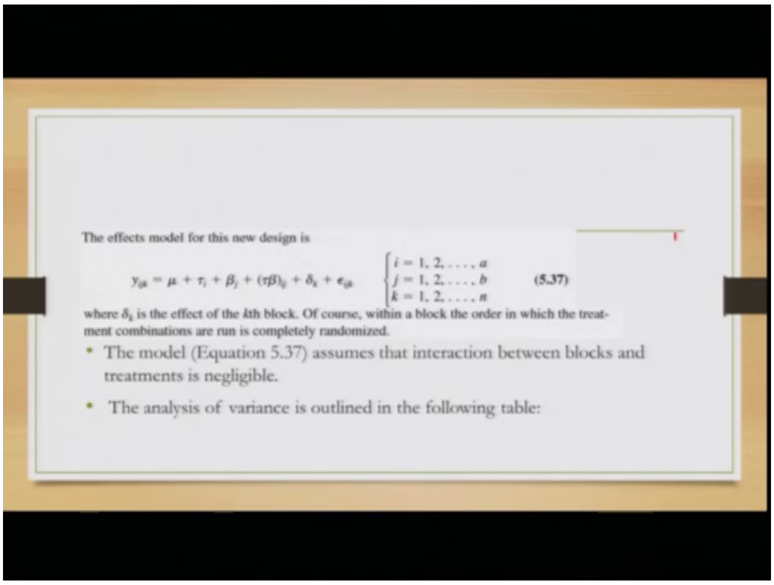


Suppose that to run this experiment a particular raw material is required; so, whatever the raw materials would be depending on the temperature variations. So, this raw material is available in batches that are no large not large enough to allow all of these treatments on how many number treatments are there? There are a number for factor A, b number for factor B, and small n for the total number of samples. So, the total number of so called set of observations we are going to take to do the study would be small a into small b into small n as noted down here which am just highlighting.

So, consider the requirement is not large enough to allow all a, b, n treatment combinations to be done from the same match. However, if a batch contains enough material for ab observations then an alternative design is to run each of these n replicates using a separate batch or raw materials in order to do the experiment. So, basically consider the raw material dv variations may come out if you do not have the whole lot of raw materials to conduct the experiment considering there are a into b into n number of observations.

Consequently, consequently, the batches or raw material represent a randomized randomization restriction or for or the block. And a single replicate of a complete factorial experiment is re run within each block in order to basically understand what are the effects.

(Refer Slide Time: 15:33)



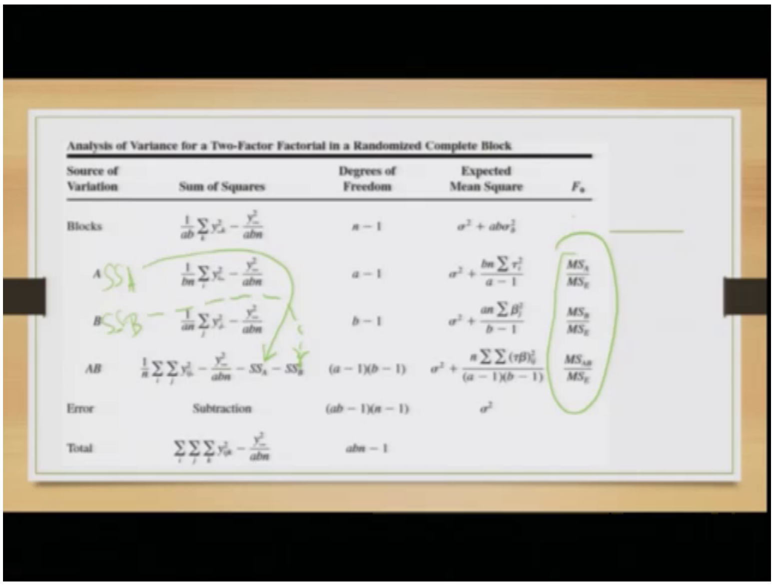
So, now as I am basically considering the effects of the of the blocks, the model now intuitively changes in this way. Now, listen to me carefully, you will understand. The effects are we want to find out on y ijk which is there, which is there on the left hand side where I am pointing. Now, on the right hand side, we want to find out the factors or so called effects coming out of the factors, so that again remains mu which is basically average of the average of the average, tau suffix I remains beta suffix j remains tau beta suffix ij remains the error term elements.

Now, we are also considering the effect for the kth block. So, if we have basically divided into blocks, so we will consider some delta based on which the effects of the blocks can be found out. In case say for example, the overall raw materials for this example is changing and we are taking different blocks together for different raw materials. The raw material type remains same, but we are using raw materials at different intervals of time from different sources. So, they may be variations in between.

So, which delta is the effect of the key block. So, k as you know is basically from one to n, you have divided it accordingly. Of course, within the block the order in which the treatment combinations are run are completely randomized, but outside. So, if you take k blocks, but in each block, they are randomized, but the different blocks have been not been ran randomized because if you remember we are taking or consuming raw materials accordingly as they come. So, they are not under our control.

The model which is given here assumes that interaction between blocks and treatments are negligible. So, if we consider the blocks and the treatments their effects or the so called in a very simple terms the correlations as 0. The analysis of variance is outlined in the following table. So, what are these and I basically go through them separate one by one. The table conceptually remains the same. The first column is basically write down all the factors.

(Refer Slide Time: 17:52)


Analysis of Variance for a Two-Factor Factorial in a Randomized Complete Block

Source of Variation	Sum of Squares	Degrees of Freedom	Expected Mean Square	F_0
Blocks	$\frac{1}{ab}\sum_k y_{..k}^2 - \frac{y_{...}^2}{abn}$	$n-1$	$\sigma^2 + ab\sigma_\delta^2$	
A	$\frac{1}{bn}\sum_i y_{i..}^2 - \frac{y_{...}^2}{abn}$	$a-1$	$\sigma^2 + \frac{bn\sum \tau_i^2}{a-1}$	$\frac{MS_A}{MS_E}$
B	$\frac{1}{an}\sum_j y_{.j.}^2 - \frac{y_{...}^2}{abn}$	$b-1$	$\sigma^2 + \frac{an\sum \beta_j^2}{b-1}$	$\frac{MS_B}{MS_E}$
AB	$\frac{1}{n}\sum_i\sum_j y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B$	$(a-1)(b-1)$	$\sigma^2 + \frac{n\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
Error	Subtraction	$(ab-1)(n-1)$	σ^2	
Total	$\sum_i\sum_j\sum_k y_{ijk}^2 - \frac{y_{...}^2}{abn}$	$abn-1$		

So, here if you note down, they would be for the blocks, they would be basic blocks means the k number of blocks you considered, they would be effort coming from factor A, effort coming from factor B, effect coming from AB combined. They would also be a error, and com combination of all these things would be told total error. The second column I am only reading the column heading.

Second column would be the sum of the squares; third column obviously, would be the degrees of freedom; fourth column, obviously, very intuitively we can now say it would basically be the mean square error or the expected mean square values. Mean square would basically mean depending on whether you have all the observations for the whole population; if you do not have you want to basically find out the expected value of that. And then; obviously, it will be the f statistic value which will basically be the ratios or the mean square a which you have.

Now, if we pay attention, so I will only pat at in special attention to the values individual cell values in this table corresponding to the second column which is basically the total sum of the squares. Now, if I consider the total sum of the squares, the blocks basically would be in that you have kept the blocks fixed. And for each block you have summed them up for all the factors A and all the factors of B which means it will be y square and the suffix you note down very carefully will be dot because you are summing up for all the i‘s second dot because you are summing up for all the j s. But the k remains because the blocks are different.

So, as mentioned here it will be summation for all the case y squared dot k and; obviously, if you want to find out the averages it will be divided by AB. And minus value would be; obviously, as it remains is the average of the average of the average squares which is y square dot dot dot dot dot dot is for the suffix divided by abn. Similarly, when you go to factor A and factor B very intuitivel,y it would be what it will be summation y square. And If you are this factored A basically it will be I dot dot as mentioned here, I am just pointing my finger here, it will y square suffix i dot dot. And outside it will be divided by b into n as it is. And the minus value remains as it is if it is y squared dot dot dot divided by abn.

And if you go to the factor B, it will y square dot j dot as mentioned here divided by because you are summing up on all the j s which is it will be divided by a into n as mentioned. And the minus tau would remains as it is which is y squared suffix dot dot dot divided by abn. Now, when you go to the the factors of ab, so these values are. So, these are basically SS A and SS B. So, SS A is here. And let me change the color or let me put it making this way SS B.

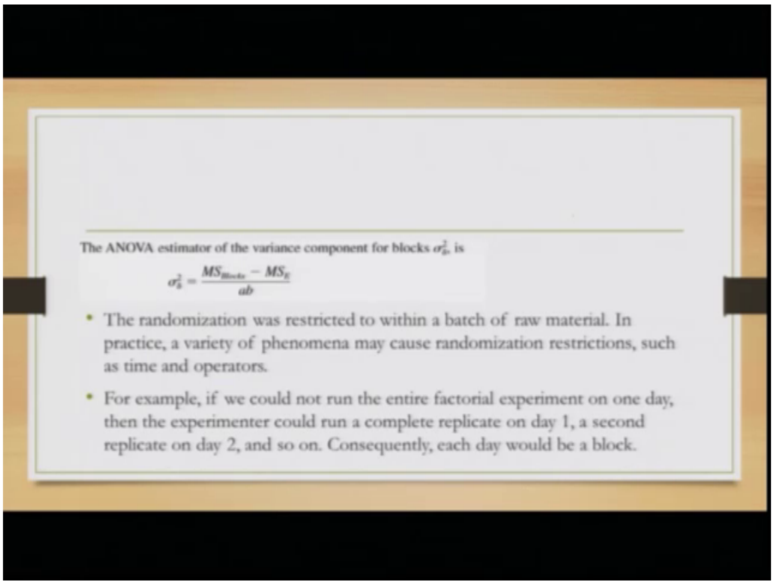
So, now if you want to find out for a b, so a b would basically be kept fixed for i and j it will be y square suffix ij dot divided. And the whole term summed up for all the i and j s and it will be divided by n. And minus term would obviously, always remains and the minus which is y square suffix dot dot dot divided by abn, but here all as you are finding on the next level the formulation of ab, it will be minus of SS A and minus of SS which you have already noted.

What they are then we find out the error term by the subtracting, and the total value remains as y square suffix basically i, j, k. Because as you are finding it for all the values of i, j, k minus term remains as it is which is y square for all the dot remains for all the suffixes divided by abn. Similarly, the degrees of freedom are found out, now, again even though I have not been discussing the degrees of freedom I will still try to highlight it once again, but this is very intuitive and very simple.

For the blocks, as they are n number of blocks it will n minus 1. For factor A as there are a number of factors it will a minus 1; for b as there b number factors, it will b minus 1; for ab it would basically be a minus 1 into b minus 1, and the errors would basically be simply very simple like this. It is a b minus 1 into n minus 1. And it can basically do the calculations.

And if you remember degrees of freedom total or intuitively obviously, remains abn minus 1. Based on that you find out the expected mean square, this expected mean square gives you the values based on which you will calculate f naught and this f naught values are calculated where am just highlighting. So, once you find out you can either basically reject or accept the null hypothesis.

(Refer Slide Time: 22:51)

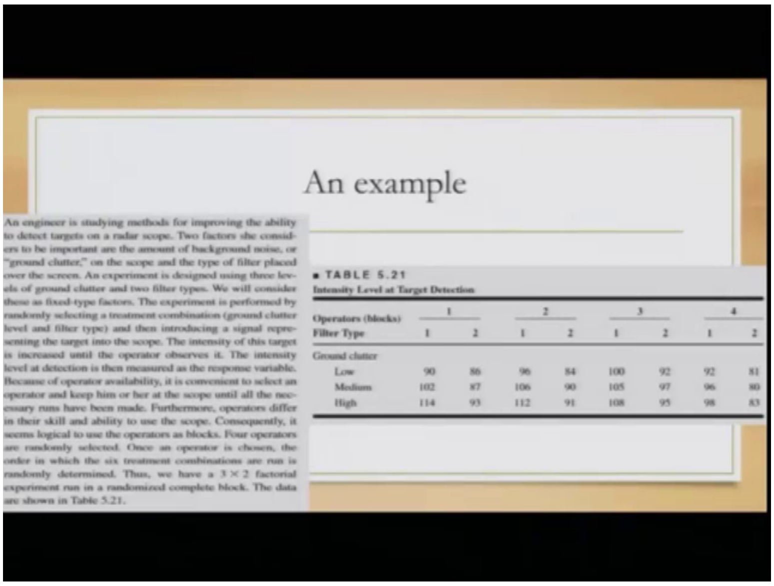


The ANOVA estimate of the variance component the blocks which is basically corresponding to sigma square suffix delta for the fact that delta is basically the so called dispersion we are taking for the blocks. It will be in the mean square of the blocks minus mean square of errors divided by a into b, because you have the sum of for all the factors number a and all factor number b. The randomization were restricted to within a batch of raw materials in practice a variety of phenomena may cause randomization restrictions as just time and operations or raw materials quality being different from different batches.

So, say for example, I get a two different batches one from vendor a and one from vendor b, and we want to maintain the quality. So, if it was only for vendor a, then those concept of randomization would have been taking place, but if they are not then these concepts can be utilized in order to find out what is the effect. For example, if we could not run the entire factorial example experiment in one day, then the experimenter could run a complete replicate on day one a second replica and day two and so on and so forth. So, consequently each day would be a block.

Say for example, time is limited and he has to do the experiment. 8 hour shift is there. In the 8 hour shift, say for example, he does it for free number of blocks and that rest day meaning two he will do in the next day which is day 2. And then they can because the formulas as I considered can be or we have discussed in the last few minutes could be considered in order to basically proceed with the example to either accept or reject the whatever h from hypothesis null hypothesis was.

(Refer Slide Time: 24:22)



So, let us consider an example and engineering studying methods for improving the ability to detect target on a radar scope. Two factors she considers to be important the amount of background noise. So, what are the amount of background noise which is there or the ground clutter or the scope of the type of filter which is being placed on the screen. An experiment is designed using three levels the ground clutter and three levels of ground clutter. So, depending on the effects and they basically there are two type of filters for the noise filters. So, we consider there these are the fixed type factors the experiment is performed by randomly selected treatment combination using the ground clutters, and then introducing a single representation of the target on the scope.

So; obviously, it will have an effect both from the ground clutters and the filter types. The intensity of this target is increase until the operator observes it. So, you are trying to basically increase the intensity in order to observe it on the screen. And because the operator availability is convenient to select an operator and keep him or her at the scope until all the necessary trance and the readings have been done. Furthermore, operators differ in this skill and ability to detect. So, say for example, I am able to detect at a certain level of intensity, someone else is basically able to detect it at a higher or lower level of intensity, so obviously, those have to be considered in when you are trying to do this experiment.

Consequently, it seems logical to use the operators as blocks. So, if there are three operators I will consider operator one as block one, operator two as block two, operate a three as block three. So, the number of experiments which you are going to do I am going to divide it accordingly that is very simple.

Once an operator is chosen the order in which the six treatments combines or run at randomly and determined. So, it can be operator two loading first then operator one then operator three or it can be operator three operator two operator one. So, whatever consequence we have or the sequence we have. Based on that the data is shown in the table which is ground clutter which is low medium high and you have basically the auto operator levels divided into filters of type one, type two, type three and type four.

So, with this I will I will basically go into the details of the value. So, under operative block one, here basically 90 to 114; for operator two it is basically for filter type it is basically 86 to 93. Similarly, for two three four the values are given now based on that you can just plug in the values and which you will do in the next class and try to basically understand how the problem is solved.

With this, I will end this 22nd lecture and continue discussing this problem with the factors blocks are considered in more detail, so that it becomes very easy for us to understand later on.

Thank you very much and have a nice day.