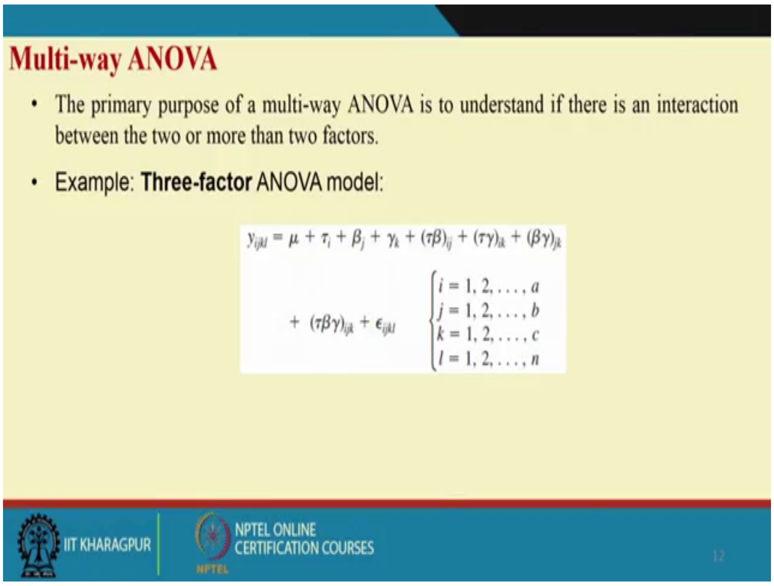
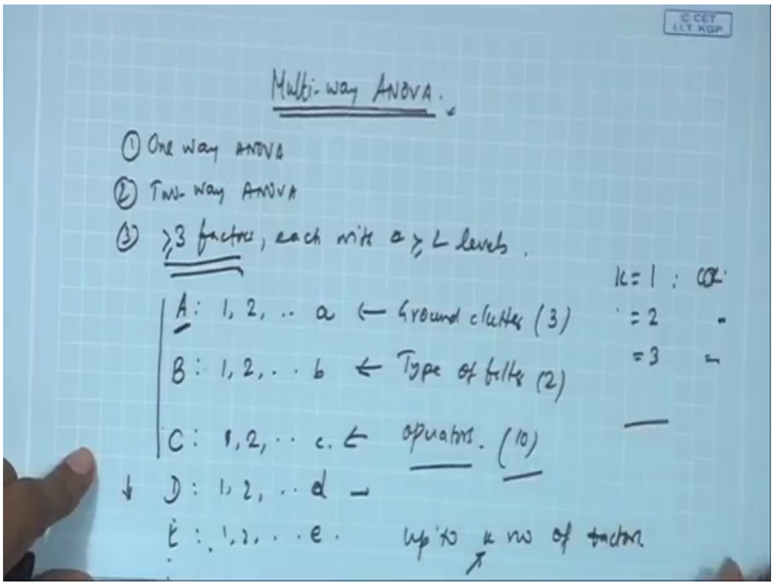
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 18
Multi - Way ANOVA

(Refer Slide Time: 00:20)



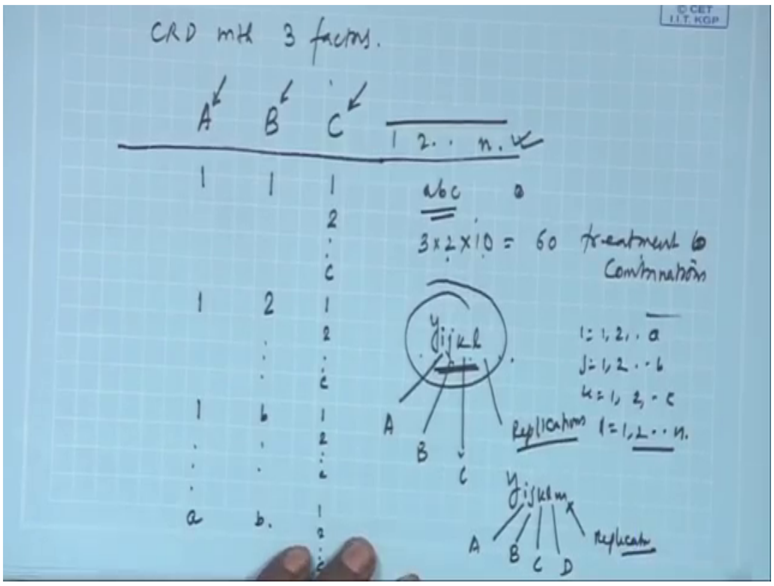
(Refer Slide Time: 00:22)



Welcome. Now we will discuss multi way ANOVA. I hope that you have understood one way ANOVA, two way ANOVA. So, now we will be discussing multi way. Multi way means 3 or greater than equal to 3 factors, where each with greater than or equal to two levels. For example, suppose you consider 3 factors A B and C, A with a number of levels, B with b number of levels, C with c number of levels. Then, this is 3 factor model and you, it maybe something like, in our radar scope example, this one is our ground clutter, a is ground clutter, b is type of filter, c can be anything, some other factor. For example, if you have skill operator plenty in number.

So, you can also control the operator. So, that mean the operators can be controllable at that, in that case. So, then the 3 different level, every 3 different factors, each with different levels, ground clutter 3 level, this is two level, this may be suppose you have 10 different factors, operators, who can be considered, you have available this. So, this is 3 level. Now what happened you may find that there are more number of factors d, e many more. So, then D will be 1 2 at each with d levels, E will be 1 2 each with e levels, similarly you can go up to k number of factors, k can be any value, if k equal to 1, this is 1 factor c experiment, k equal to 2, two factor experiment, k equal to 3, 3 factors experiment, k equal to 4, 4 factor, k equal to k factor experiment. Now, when we are talking about multi way ANOVA, you please understand that we are basically considering an experiment with multiple factors which means 3 or more number of factors. So, there is a complete randomized design CRD.

(Refer Slide Time: 03:42)



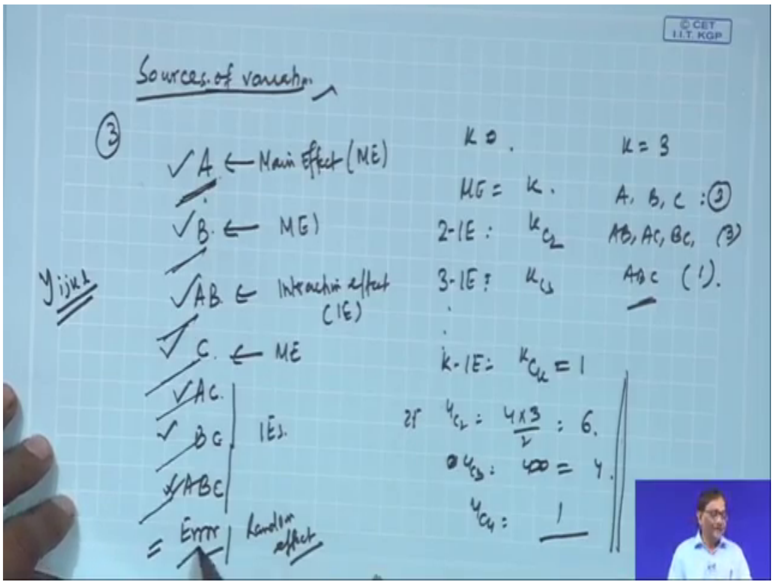
Let CRD with 3 factors. So, what will be your data set? I mean how you will know what will be your experimental settings. So, A B C.

Now, A level 1, B level 1, B can be level 1 2 like c level, a level 1 b level 2 c can be 1 2 c levels. Similarly what happened again a dot, dot, dot a level 1 and then b level b, c will be 1 2 c in this manner a level, then finally b, b level and c will be 1 2 c levels. So, how many experimental settings will be there a b c, if a equal to 3, b equal 2, c equal to 10, then you have 10 into 20 60 treatment levels, treatment combinations, not level treatment combinations.

Suppose you have done 60 treatment experiment with 60 treatment combination, and again, each level, each and each combination, you have 1 2 like n number of observations. So, essentially, then you have i in ith index for factor a, jth index for factor b, kth index for factor c, and l if lth index for number of replications. So, essentially your number of a general observations will be represented like y i j k l. This is i for factor a j for factor b, k for factor c, l for replications. So, this is the general observation when you conduct experiment complete.

Randomized experiment, you will be getting random response values y i j k l, where i varies from 1 to a, j varies from 1 to b, i varies from 1 to a, j varies from 1 to b, k varies from 1 to c, l varies from 1 to n. So, if there is a one more factor d, then what will happen your general observation y i j k l m. So, i for a, j for b, k for c, l for d and m for replications . So, when you increase the number of factors. So, what is happening your sources of variation is increasing source or sources of variation.

(Refer Slide Time: 07:00)

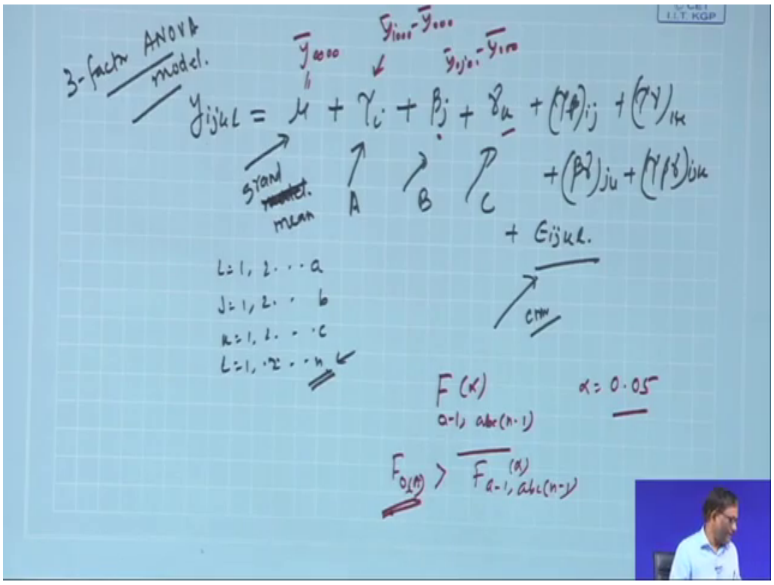


So, it is analysis of variance and sources of variance, are very important. So, what will be for 3 factor case, what will be the sources A, B, then their interaction, then C interaction AC, BC, another interaction plus error. What is happening then, you are having A factor which is known as mean effect, B interaction is another mean effect, mean effect ME, AB is interaction effect (IE), C is another mean effect, ABAC these are all and ABC these are all interaction effects and this is the random effect error. So, depending on the number of factors K depending on the K, you will be having different kind of effects. So, if you experiment with K factors, then your main effect will be K numbers interaction effect will be of different kind, plus this two interaction two way interaction.

So, this will be KC two numbers, 3 way interaction, this will be KC 3 number like this K way interaction that will be KCK means 1. So, if I consider K equal to 3 in that case, mean interaction is ABC; that is 3 numbers two way interaction is AB AC BC, again 3 numbers and 3 way or K interaction is ABC; that is 1. Now C 3 3 is 1. So, how many sources you are getting 3 mains effect source, 3 inter two way interaction source, another one is that A, B, C that 3 way interactions. So, ultimately 1 2 3 4 5 6 7 8, another one is error is there. So, essentially by sources of variation in, we are talking about the entire the source the variation of that you will observed in case of 3 factor experiment, yijkl data if you see that there is a total variation in the data some data, data will be vary, value will vary and there will be variation.

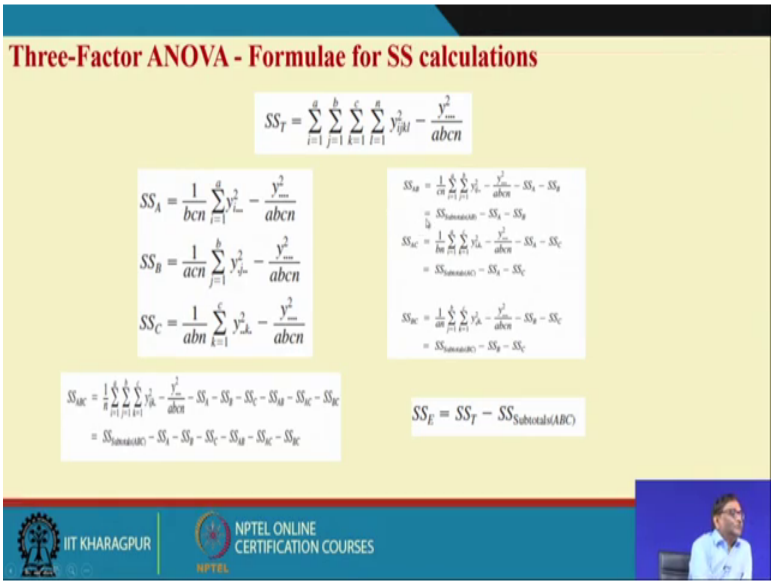
And as a result this total variation can be decomposed into so many variation level. So, many parts for A for B for C for AB AC BC ABC and error. Now, if you go for 4 then what will happen? Your interaction effect will be 4 c 2 means 4 into 3 by 2 that will be 6; that is two way interaction will be 6, 3 way interaction will be 4 c 3 4 by, what will be the 4 c 3 means 4, then 4 c 4 means 1. So, in that way you have to find out the interaction accordingly, you find out the sources. You may argue that there are 2 factors. So, 3 factors or 3 source and on a random source fine, but in a multi way analysis of variance or even in analysis of variance gives that a mean effect or the source, but their interaction it can also considered that the source because if there is no interaction if you can the change will be there. So, accordingly we will partition the things. So, a general observation yijkl, this will be partitioned into what it will be there will be grand mean or grand average. There will be A effect, there will be B effect, there will be AB effect, there will be C effect, AC effect, BC effect, ABC effect and error will be there, that is what we will do

(Refer Slide Time: 11:47)



And so, accordingly can what you will write yijkl, this general observation. You can write this equal to mu plus tau i mu i is grand mean this is the overall population model. So, grand mean tau i for factor a, beta j for factor b, plus gamma k for factor c plus tau beta i j plus tau gamma i k plus beta gamma j k plus tau beta gamma i j k plus epsilon i j k l . This is your 3 factor ANOVA model. So, I saw 1 2 a j varies from 1 to b, k varies from 1 to c, l varies from 1 to n. So, we are considering valance ANOVA, because sample values equal and it is better to consider a valent and this is your error tau, now if you want the estimate things what will happen, what are the estimated value mu, estimate will be 4 see this we will go for y dot dot 4 that bar, this one will be y i dot dot dot bar minus, this similarly for this that mean, what will happen i dot j dot dot bar minus y dot dot dot this, k will come and ultimately interaction effects you have to find out. So, that t is similar manner the way we have done for, two factor ANOVA or two way ANOVA case. Now let us see the see the calculation part.

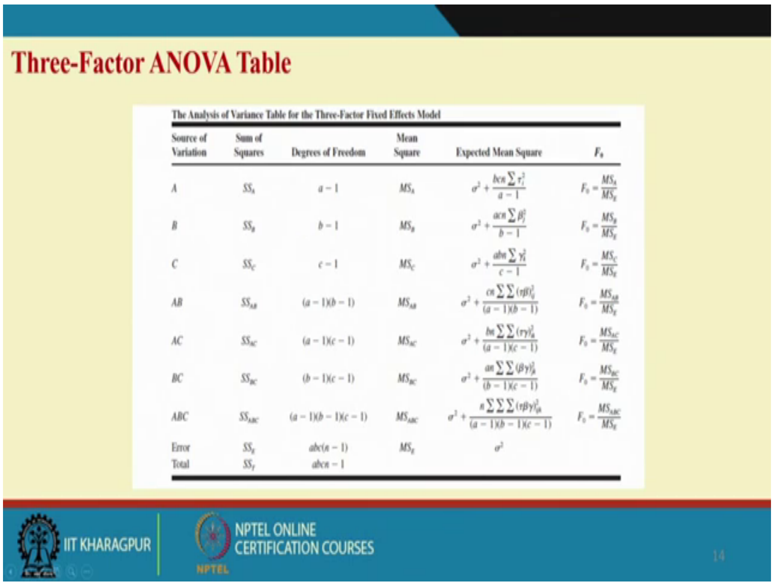
(Refer Slide Time: 14:03)



So, you see essence. So, what do you want you want to compute SST. I am not going into the further partitioning effect because that the concept remains same, you know all those things. So, let us see that what is SST that mean you are talking about cos squaring each of the observations and then taking sum and we were subtracting the subtracted factor this is nothing, but again the total square grand, total square by the number of observations then you find out SSA SSB SSC, then you find out SSAB SSAC SSBC, when you are finding out SSAB, you require SS for subtotal for AB when you find out SSAC, you require SS subtotal for SE and similarly for BC for SS subtotal for BC and this is formula and once you have all.

Those things as well as your SS ABC then SS ABC subtotal and this then you can find out find out SS E that SST minus SS subtotal ABC. So, this is just a equations are note of it is just it looks a big equation, but concept remain same like in two factor ANOVA case or two way ANOVA case, only you please you be careful about the formula on what is to be divided. For example, when you calculate SSA then BCN will be divided from the sum of the individual squares, when for B and C and will be divided. Similarly when you calculate interaction subtotal is very important that subtotal, which subtotal is relevant for which kind of interaction that you have require, to be you are going to be understood and through practice it is possible.

(Refer Slide Time: 16:18)



The Analysis of Variance Table for the Three-Factor Fixed Effects Model

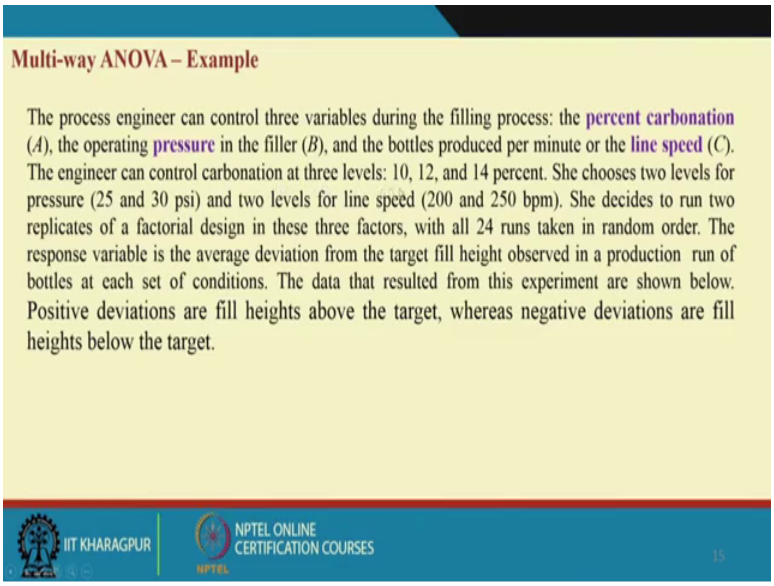
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square	F_0
A	SS_A	$a-1$	MS_A	$\sigma^2 + \frac{bcn\sum \tau_i^2}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2 + \frac{acn\sum \beta_j^2}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2 + \frac{abn\sum \gamma_k^2}{c-1}$	$F_0 = \frac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \frac{cn\sum\sum (\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \frac{bn\sum\sum (\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$F_0 = \frac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \frac{an\sum\sum (\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$F_0 = \frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \frac{n\sum\sum\sum (\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$F_0 = \frac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	σ^2	
Total	SS_T	$abcn-1$			

So, then you will definitely prepare this kind of ANOVA table and again you see that all A, B, C, AB, AC, BC, ABC, error and total I have discussed this, so variations. Then you find out sum square SSA, SSB, SSC, SSAB, SSAC, SSBC, SSABC like this, then degrees of freedom n minus 1, b minus 1, c minus 1, a minus 1 into b minus 1, a minus 1 into c minus 1, b minus c minus 1 like this ANOVA then abc into n minus 1 and abc n minus 1 divide SS by their respective degrees of freedom you get mean squares and here we are introducing one more concept called expected mean squares from theory you know that, it will be like this and although I say you know, but you we have never discussed, these please keep in mind that it is expected that the so, every source of variation has some effect in that case, this is the quantity bcn tau i square by a minus 1 or a c and beta j square this is the quantity which is added 2 sigma square because sigma square is the quantity which is the random variation and which is related to primarily related to error, you see that SSE and MSE calculate and then the expected mean square that mean expected value of MSE will be sigma square.

So, this interesting thing that is also in the earlier cases we can into do the same thing, but here is interesting thing is that when you are calculating f zero you are basically comparing MSA with MSE, now that mean the expected value of MSE is sigma square, if there is no significant difference of the different treatment levels, for suppose for factor a or b or c, then this will concede or will be all most equal to MSE; that means, this quantity bcn see tau i square by m minus n or sen; this added quantity. This will become 0. So, that mean it the variability is or equal to sigma square which is equal to MSE, if there is difference in that expected mean square, then these each of the sources are contributing towards the difference that is what is interesting to you.

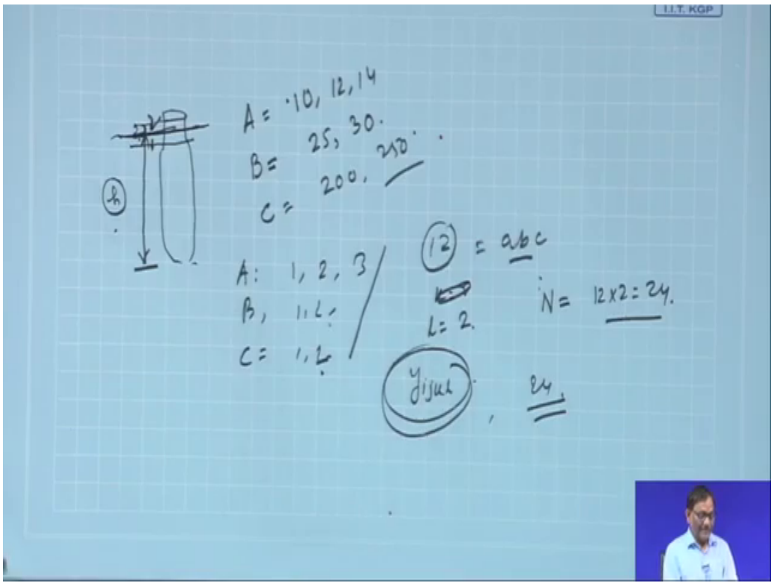
Anyhow, suppose you have only two factor case a b and a and a b and error in that case that time only this bc and a that terminology is suppose b c and a here b c is coming because it is a 3 factor case bc bn will be there c will not be there, but in the same manner this can be computed . Now there will be tabulated value for n for a tabulated value will be f what will be the tabulated value for a, f alpha a minus 1 and in this case the what is the error degrees of freedom a b c into n minus 1 a usually alpha equal to 0.05. Now if your computed value for a this is greater than this tabulated value a minus 1 abc into a, n minus 1 alpha, then this has l have effect. Similarly b similarly c similarly like this.

(Refer Slide Time: 20:17)



You see one example, the process engineer can control 3 variables during filling process, the percent carbonation, which is denoted by a, the operating pressure in the filler denoted by b and the bottle produced per minute or the line fee speed denoted by c. Actually this is basically that coke making coke in the Coco-Cola or something like this the filling the bottle. Now the engineer can control carbonation at 3 levels 10, 12 and 14 percent, c chooses 2 levels for pressure that 20 5 and 30 psi and 2 level line speed 200 and 250 bpm bottles per minute, c decides to run two replicates of a factorial design, in these 3 factors with all 24 runs to be taken in random order. The response variable is the average deviation from the target feel h8 observed in a production run of bottles at each set of conditions. The data that resulted from this experiment are shown in the next slide, positive deviations are feel h8s above the target whereas, negative deviations are feel at below the target. What is happening, here by situation is like this.

(Refer Slide Time: 21:53)

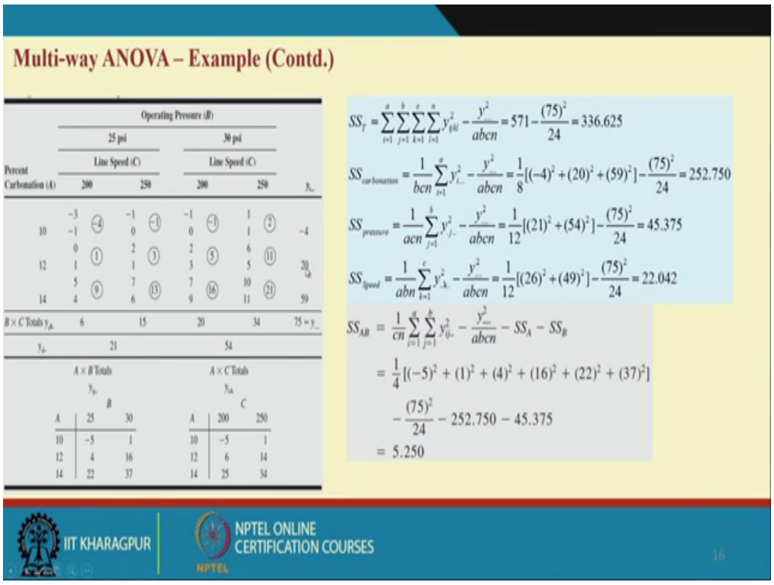


Suppose you have coke bottle. So, you want to fill with coke.

There are one is factor a that is carbonation, it is having 3 levels 10, 12 and 14 percent. There is factor b, which is basically the pressure at which the bottle will be that coke filled, that is two levels 25 and 30 and another one is at what speed you are basically filling the bottle that is c is 200 bottle per minute and 250 bottles per minutes. So, this is the (Refer Time: 22:33) that is line speed. Now, so, that mean what happen your a 1 2 3 levels, b 1 2 levels and c 1 2 levels. In total how many independent treatment combinations are there, 3 into 2 into 2 12 which is abc.

But the engineer decided to conduct two number of replication. So, in each setting, there will be two replication, so k equal to i j k l, l equal to 2. So, that means, your total number of runs will be 12 into 2 equal to 24, what I mean to say y i j k l this value, the total value will be how many such observation you will get you will get 20 4 such experimental data.

(Refer Slide Time: 23:45)



Now see this, here you see this one this complete this is the tabular representation, please remember again I repeat we have taken from Montogomery book percent carbonation 10, 12, 14 your operating pressure 25-30,

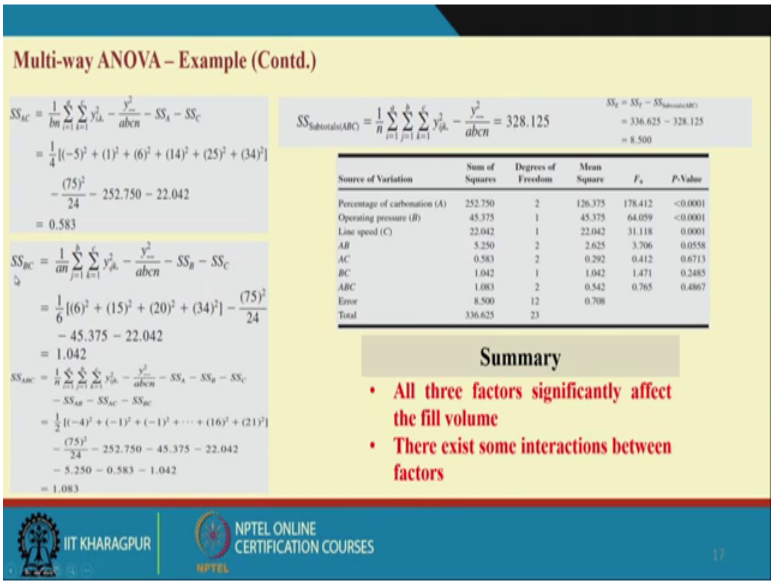
and then line speed 200 and 250. So, how many observations are there, 3 into 2, 6 observations and again replication 2 is cases of 12 observations here, 12 observations here 20 4 observations. You see the right hand side SST calculation that mean you find out the square of each of the 20 4 y values, y i j k l square. So, this value once you and then your subtracting by the grand total value which is when I summing up all those thing this is 75. You may be a may be thinking that what is this minus and plus basically this is nothing, but suppose you want to feel up to this level or up to this level let it be.

So, this is suppose h h8 this to be field now in bottle 1 you measure and you find out that this is a gap bottle to you measured maybe this is the gap suppose this is the target let it this is the target which is h and this is positive more than target this is negative less than target it is there. So, in that manner 20 4 experiments are done and in each set in two bottles, two experiments are a two replications are there, like in the first case it is minus 3 minus 1, second case may be zero 1 like this is the field to the target. So, response variable y this is basically deviation from target filling, target is h this much will be total you fill and if it is less that or more then that will be represented by y.

So, this formula now SS carbonation means basically SSA, this is the equation you have seen earlier and then what you are doing 1 by bcn y i is changing what is i here i is basically this first row this related to 10, 12, 14 and you see what you are writing here, minus 4 square plus 20 square plus fifty 9 square. What is this? This is basically the row total. So, in the first row related to carbonation that is 10 here every where two replications are there and ultimately the total is minus 4. So, minus 4 square, second total is 20 and like this and you divided by bcn, b is 2, c is 2 and n is 2. So, 1 by 8 you are getting like this one similarly find out.

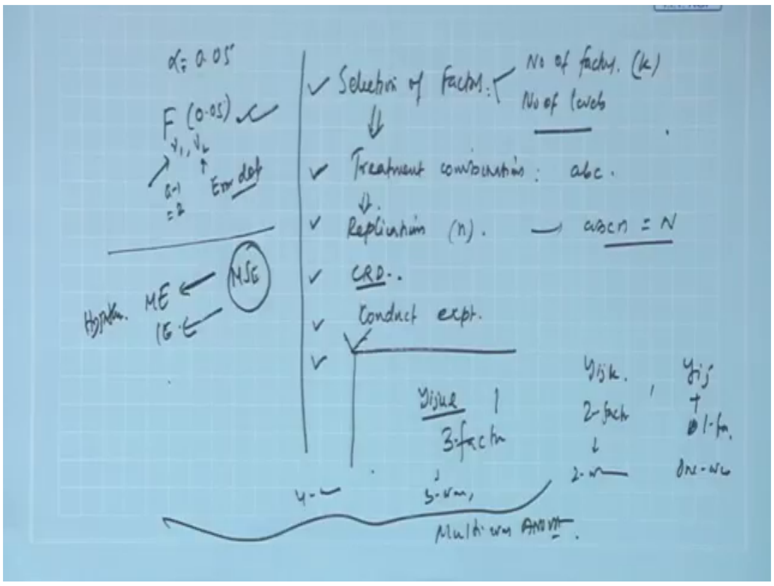
SS pressure, SS speed and then SS AB which is basically the sub total. This is the sub total 1 by cn, i j square and then i j. This one this is the sub total and these are the SS pressure you will be subtract it, now subtotal y i j i stands for a for a i and for j, it is basically for j is operating pressure of b then dot, dot mean the total. So, that mean 20 5 psi 10. So, how many observations are there. So, how many observations are there 1 2 3 4 observations are there. So, what is the total of these observation here this total is minus 4 this total is minus 1 minus 4 minus 1 that is minus 5.

(Refer Slide Time: 28:01)



So, in this manner you compute and get this one and similarly for other values also SSAC SSBC SSABC and like this and finally, and SS sub total abc also you get it and then you calculate SSE SST minus SS sub total, this is the and once you have all those SS value you create ANOVA table, degrees of freedom known to you, mean square you can compute f zero ,you can compute and p here we are given p value, but it is approach each it is because p value competition is tedius.

(Refer Slide Time: 28:42)



So, what you do basically you choose alpha equal to 0.05 for some other value, but usually is alpha is 0.05 and then find out f value for these alpha given degrees of freedom nu 1 and nu 2 and nu 1 will be rest, if it is a 3 level. So, it will be a minus 1. So, that mean 2 and nu 2 always for error degrees of freedom, error dof. Here in this, case error daf is 12. So, it will be always 12 now, if this value is less than the competitive value h zero is rejected mean there is factor a effect, similarly factor b effect , c effect like this. Now see, all 3 here. What happened all few see this one that first one, second one, third one, they are basically significant because p value is much less than zero point zero 5. So, this first 3 significant if you see the every interaction every interaction here the value is .0558. This just above 0.005. So, you may say this is not significant, but if you consider that this is 0.06 is almost 0.05, you may consider also ab f a, but other probability values here p values are very high. So, this are not significant.

So, all 3 summary all 3 factors significantly affect the field volume and there exist a sum interaction between the factors. So, this is what is your 3 way ANOVA. So, you can think of that, there can be more than 3. So, I hope you have understood also 3 way ANOVA and then calculations and if not, please write in forum and , we will see we will see in the forum and our tas will be seeing the forum and they will be , be giving you the reply, but up to this I want to tell you very one very important thing; that means, that first is selection of factors and when you do experiment, you must know selection of factors, how many factors, here number of factors is very important, here number of level is also very important. So, once this is done, you are in a position you know that what will be the treatment combination this will help you in finding out the treatment combination. Suppose if I say 3 factors then your treatment combination is abc. Once you know the treatment combination then you will select that what will be the replication, number of replications. So, replications is n .This will help you to in find out that , what is the total number of experiment that you do conducted and then you just see that you cannot conduct this n number of experiments arbitrarily that will be complete randomization design.

Then what happen, after that design you have to do and then do conduct experiment. Once you conduct experiment, you get a data, you get data table. So, you will get y i j k l or if it is 3 factor if it is two factor, y i j k if it is one factor y i j. So, 3 factor case two factor case and this is one factor case. Once you this is p requisites for because these are the things you that is what you have done actually and this data set is available, now you analyze the data set if it is one factor you go for one way ANOVA, if it is two factor you go for two way ANOVA, if it is 3 factor you go for 3 way ANOVA, if it is 4 factor 4 way ANOVA and if it is 5 10 factor 10 way ANOVA . All these are known as multi way ANOVA another and everywhere because we are interested to know whether factor affects are there or not or interactions are there or not, Your hypothesis will be concerned with the factor effects, it will be mean effects hypotheses, all hypothesis related to this interaction effects and E and this actually mean effect and interaction all are compared with the error, which is MSE.

If the variability explained by the main effect is all equal if similar to MSE mean variable is similar to MSE, then there is no effect, if there is factor effect it will be much more than the MSE and then if test contabulated value, computed value, comparison and you will find out. This is what is in nutshell ANOVA and this ANOVA model has similarity with regression also.

So, we will discuss regression in subsequent classes, lectures. Thank you very much for your patient hearing.

Thanks.