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

Lecture - 38
Fitting Regression Models - III

Welcome back my dear friends a very good morning, good afternoon, good evening to all of you we are in the 38 lecture for the TQM II course under the NPTEL MOOC series and I am Raghunandan Sengupta from IME department IIT Kanpur. So, if you remember we in the last 2 lectures for the last week, which is the 36th and 37th seventh lecture we did not cover much of the slides.

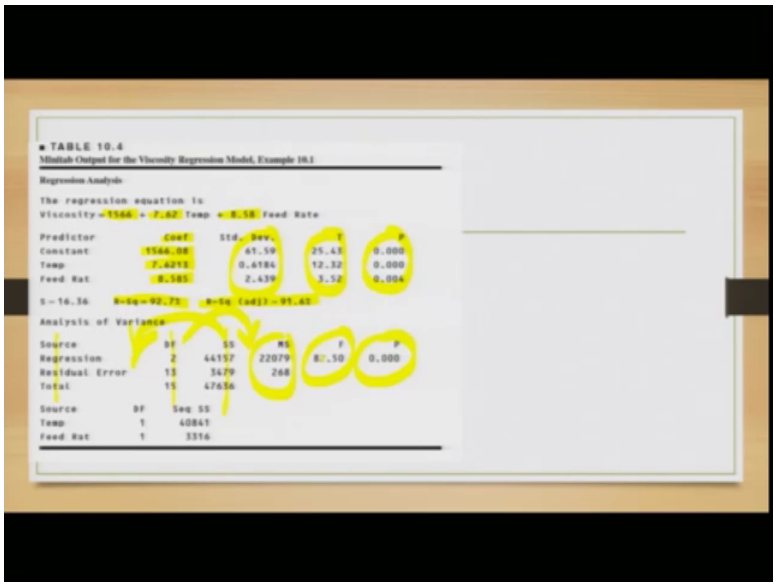
But we have we had a lot of discussions about regression model, multiple linear regression model or you find out estimate of beta how you find out the sum of the square of errors then how you find out the variance for the betas and the covariance of betas sigma square, what is the degrees of freedom. And then; obviously, I keep kept repeating what are the assumptions and how you do the q q plots, how you do the plots of the errors with respect to x 1, x 2, x 3 whichever variable which are there the assumptions being that normality holds true for all the x s for the errors for the y s and expected value of the errors is 0 variances and all is obviously, there and all this things.

Then we slowly went in to the 2 to the power k fractional factor models considering there are k variables they are 2 different levels 0 1 minus 1 plus 1 and so on. So, and we saw that that how the concept of orthogonality can be brought into the picture and we can module in the multiple linear regression models.

And we were discussing that why is the variable and there are beta naught, beta 1, beta 2, beta 3 which means there is x 1, x 2, x 3 variables based on which you can proceed and we can find out the errors and the variances of beta naught variances of beta 1, beta 2, beta 3 where the variances of beta 1, beta 2, beta 3 for the last example which were discussing in the 37th lecture all the variances were equal because food for the orthogonality one, orthogonality one.

And I did mention time and again the concept of degrees of freedom you are losing the degrees of freedom as you and more and more variables which are the independent ones which basically affect the value of y. So, continuing the discussion.

(Refer Slide Time: 02:34)



So, if you have the output for the viscosity regression models at the actual viscosity which is y is given by beta naught hat which is 1566 plus beta 2 hat a beta that was beta naught hat.

So, this will be beta 2 hat with 7.62 into temperate temperature is x 1 plus 8.52 which is beta 2 hat into the feed rate and the if you see the constant. So, which is what is interesting I want to note down and highlight. So, these were the coefficients, the coefficient value is 1 point 1566.08 which is here 7.656213 which is here 8.585 which is here. The standard deviations or the standard errors for the constant one; obviously, because they are estimates, estimates would basically have expected value and if there is an expected value they are random and they would be a variance for that. So, variance for all these 3 constant which is beta naught temperatures coefficient which is beta 1

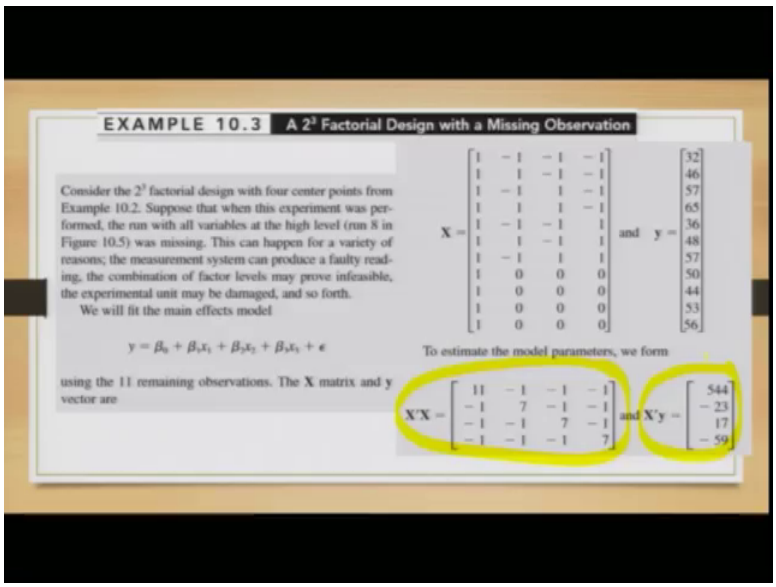
And and feed rate coefficient which is beta 2 their standard error are given. So, note down for a value of 1566 the standard error deviation is about 61 for 7.62 it is 0.61 for 8.5 it is 2.4. For the temperature difference values which you remember t plus and t minus and the p value level of significance, the r squared value is given as 92.7 adjusted

r square is given as 91.6 adjusted r square basically considers the degrees of freedom in those picture that is all, in a in a very general layman terms I am saying.

So, r square basically gives you the level of prediction we are able to do it, if it is 92.7; that means, we are able to predict about 92 percent of the variations of y with respect to x; obviously, 100 minus 92 would be the effect of white noise or the errors, the adjusted r square when we bring the degrees of freedom into the picture. So, it comes out to be about 91.6; that means, 100 minus 91.6 is the overall errors or the prediction which we are not able to predict using the axis which is basically given by the white noise.

The analysis of variance again we will go back to the general main theme of our discussions you have a table the, the reading in the all the variables of the attributes are there on the first column, then second column you have the sum of squared errors you have the third column in the degrees of freedom you have f value, because you want to test the null hypothesis whether this is true or false and you have the degrees of a level of significance. So, these are the sources, these are the degrees of freedom, these are the sum of squares in a general table this sum of the squares comes here degrees of freedom comes here, you have the mean square error which is division of the sum of squares by the degrees of freedom with a value and the level of significance. So, based on that you can calculate and do the ANOVA table for the regression model also.

(Refer Slide Time: 05:47)



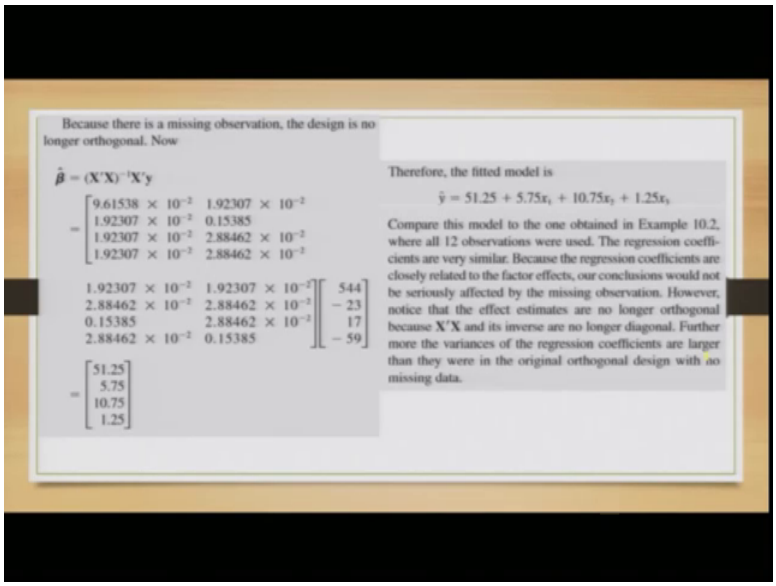
2 to the power 3 factorial, factorial design with mixing observations.

So, you are mixing the observations consider the 2, 2 the power 3 factorial design with 4 centre points from same examples suppose that when this example was performed the run with all the variables at high level run were missing. This can happen for a variety reason that measurement errors was there, measurement system can produce a faulty reading, the combination the factor level are not feasible infeasible they were not tempered or tuned as for the experiment and so on and so forth or the unit was damage, we did not get that unit or we miss the unit or the person who was taking the reading was not attentive on all this reasons can be there.

Again the same model, y is equal to beta naught, but beta one into x one plus beta 2 into x 2, plus beta 3 into x 3 plus epsilon again is the multiple linear regression model. Beta naught, beta 1, beta 2, beta 3 are the regression coefficients with respect to beta naught is when is basically the point at which where it is cutting the y axis. Beta 1, beta 2, beta 3 are the partial regression coefficients or the rate of change of y with respect to x 1, x 2, x 3 respectively. Using the 11 remaining observations the x matrix is of this form which is basically n cross the 4, 4 means beta naught beta 1, beta 2, beta 3.

So, we so if there are x 1, x 2, x 3. So, it will be n plus 4 because the fourth one is basically coming for beta naught, y values again n plus 1 the x, transpose x values is found out here which I am just circling you can use simple excel sheet or a mini tab or a mat lab to find it, then x transpose y is given from that you can find out beta hat. So, beta hat beta when I am mentioning betas or the x s of the y s are all vectors, matrices, vectors, scalar would only come for the case when you are trying to find out the sigma square.

(Refer Slide Time: 07:58)



So, beta because there is missing observation the design is a no normal orthogonal when you find it out the beta hat values comes out to be 51.25 this is beta naught hat beta 1 hat is 5.75, beta 2 hat is 10.75 beta 3 hat is 1.25 therefore, the fitted model is ah; obviously, they would not be any error term.

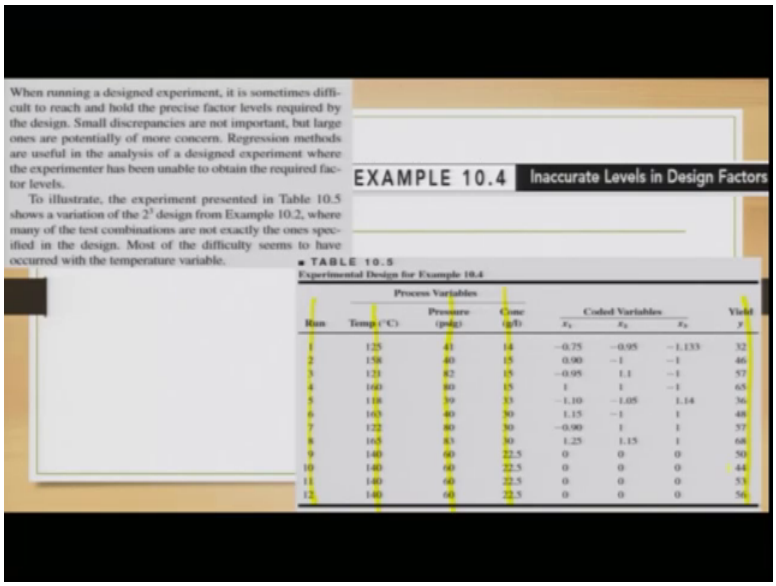
So, if you want to find out the nth plus 1 value it will be y hat nth plus 1 which is in the suffix would be 51.25 which is beta naught hat plus 5.75 which is beta 1 hat into x 1. This is the first value is the first x value and its corresponding reading is the nth plus 1, then in plus beta 2 hat which is 10.25 into x 2 and it is corresponding value is the nth plus 1 for the second variable, plus beta 3 hat which is 1.25 multiplied by x 3 and it is corresponding x 3 value being nth plus 1.

So, from that we can find out y hat, then find of the difference between y and y hat for the nth plus 1 nth plus 2 nth plus 3, n is basically depending on at which stage you are which reading you are, find out the errors plot the errors with respect to x 1, x 2, x 3 find out the averages values come out to be 0 yes or no. Then plot the values of errors which should be normally distributed with respect to the standard normal distribution you use the q q plots, plot it along the 45 degrees line check whether they are there. Obviously, they would not be all lying on the 45 degrees line, but you can get an answer for that, compare this model to the obtained earlier the regression coefficients are very similar

because the regression coefficients are closely related to the factors effects. So, what is the effect of the x 1 and y, what is the effect of x 2 and y, what is effect of x 3 and y?

So, they would be given by beta 1, beta 2 beta 3, but we are using the small sample whatever it is to predict the values using the estimate beta 1 hat, beta 2 hat, beta 3 hat, obviously, we will use beta naught hat to find out beta naught. However, notice that the effects estimates are no longer orthogonal because x transpose x and it is inverse on no longer orthogonal for the more variances of the regression coefficients are larger than then when it was the original data because you have basically shrunk your sample size to smaller simple size.

(Refer Slide Time: 10:21)

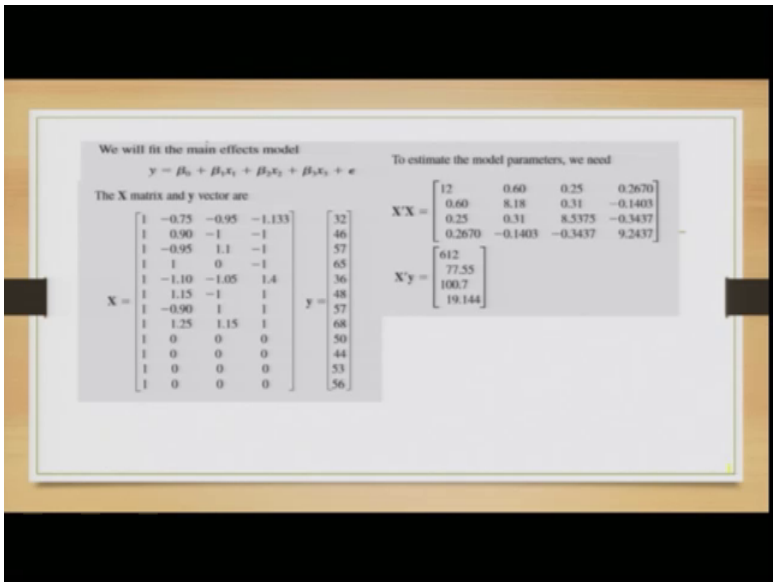


When running a design experiment is sometimes difficult to reach and hold the precise factor levels required by the design, small description are not important.

So; obviously, they would be, but large once are potentially of more concern to us and we should be careful, regression methods are useful in the analysis of the designed experiments. Where the experimental has been unable to obtain the required factor levels at what factor they are, to illustrate example where very many of the test combinations are not exactly the one specified the in design most of the difficulty seems to have been occurred with the temperature variable.

So, again I am noting down or presenting to you the experimental design for the example where you have the run given in the first column, the temperature which is the x 1 variable at the second, pressure which is x 2 in the third and concentration which is x 3 on the fourth, the coded variables are basically done accordingly such that you can find out the average values and noted down and the y yield which is the dependent variable which is given on the last column.

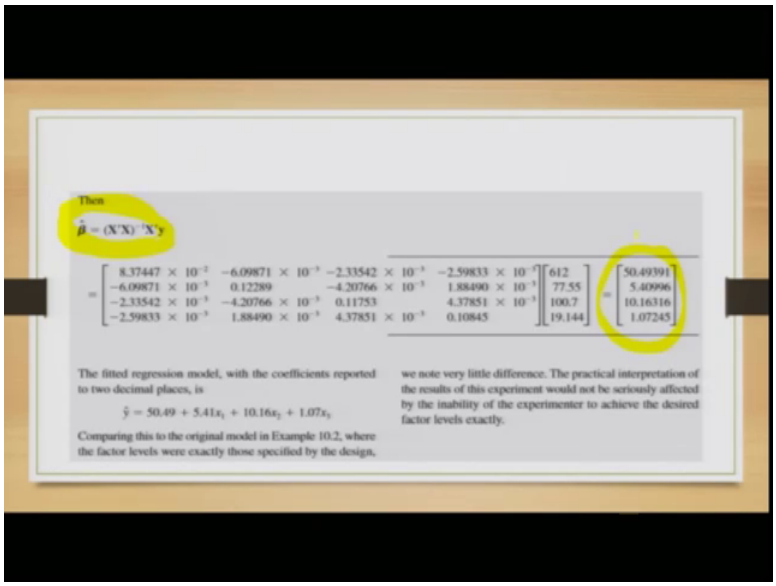
(Refer Slide Time: 11:29)



We will fit the main effect model again the same model y is equal to beta naught plus beta 1 into x 1 plus beta 2 into x 2, plus beta 3 into x 3 plus epsilon and y hat would basically be is equal to beta not hat, plus beta 1 hat into x 1 plus beta 2 hat into x 2 plus beta 3 hat into x 3 plus; obviously, there is no error.

So, x matrix is given the y matrix is given from this we find out the x transpose x and x transpose y from that we find out the beta hat vector.

(Refer Slide Time: 12:03)



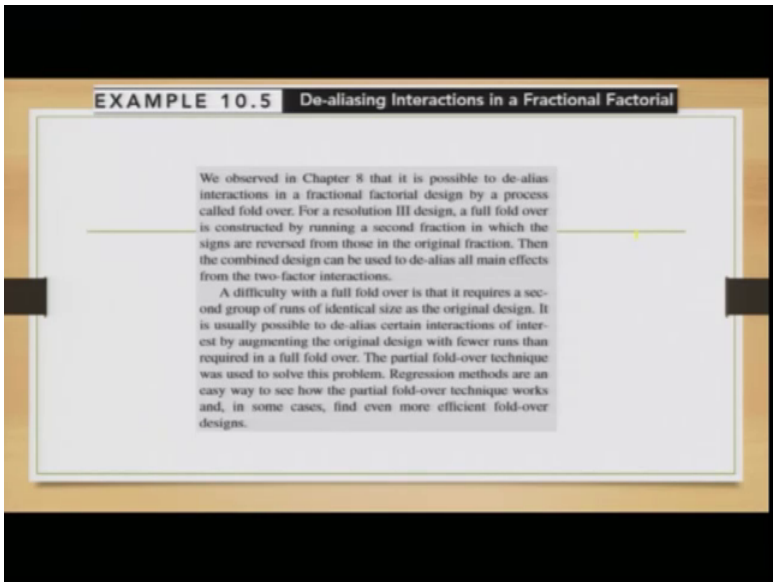
So, vector comes out to be, after the all the calculations because the formula or which I am circling always remain the same you do not have to worry ones, actually the basic concept is sum of the squares difference partially differentiate with respect to betas put it to 0 those are the orthogonal equations find out the hats and after that your job is done you proceed with the calculations.

The values of beta naught is given as beta naught hat is 50.49 I am only reading till the second place of decimal beta 1 hat is 5.40, beta 2 hat is 10.16, beta 3 hat is equal to 1.07. The fitted regression model now would be y hat is equal to 50.49 plus 5.40 into x 1 plus 10.16 into x 2, plus 1.07 in to x 3. So, this is what is given, comparing this is with the original model where the factor levels were exactly those specified by the design.

We note a very very little difference, the practical interpretation of the results of the experiment would not be seriously affected by this, hence we are able to utilise this model convincingly. So, even you even if there is inability on the experimental side like missing get was there or the machine was not working to measure rate of the temperature difference were there still there is no problem, because errors are very small errors with respect to the betas and the error terms. And; obviously, we would difilt for all these things I am, I am skipping it for the last example we will draw the q q plots draw the error term on the y axis with be respect to x 1, x 2, x 3 and do all the comparison to find out the mean of the errors is 0.

The variability of the errors should be norm a sigma square hat which want to find out and the normality conditions would also hold true we will also find out the covariance of the beta vector. So, where the principle diagram would be the square of the standard error of the betas and of the diagonal element with the covariance existing between the betas.

(Refer Slide Time: 14:09)



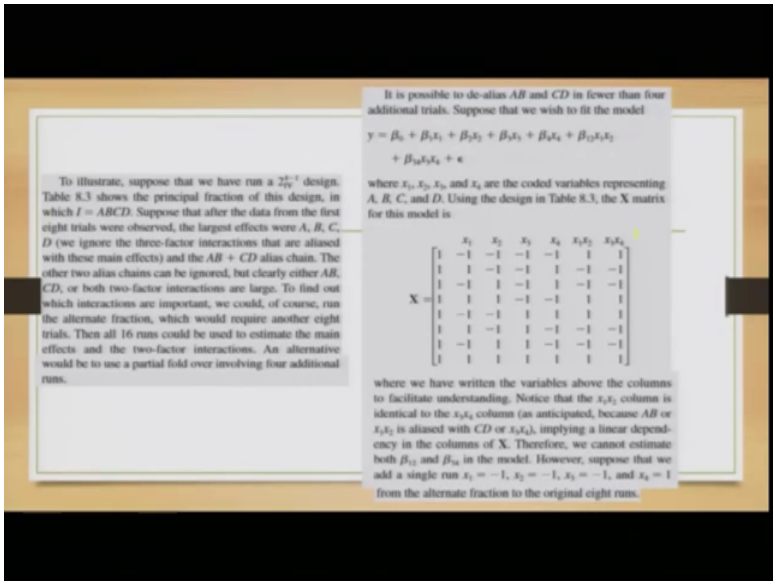
We observed that is possible to de alias the interaction in the fractional factorial design by a process called folded over.

So, you fold it and try to basically consider in using blocks, block design also for resolution of 3 degrees or 4 degrees a full fold over is constructed by running a second factorial in which the signs are reversed from those in original factors. Then the combined design can be used to de alias all main effects from the 2 factor interactions and basically have a better picture on how the effects are being done. Now, a difficulty was the, with a full fold over is that it requires a second group of runs of identical sizes.

But there that data size may not be applicable because the number of observations which we have for the sample is low, it is usually possible to de alias certain interaction of interest by augmenting the original design with fewer runs than required in a full fold over, the partial fold over technique was used to solve this problem because we would not to do the folds in at one go for considering all of the observations at 1 go, we need to partially break into samples and then do the folds.

The regression methods are an easy way to see how the partial fold over technique works and in some cases find even more efficient fold over design methods.

(Refer Slide Time: 15:25)



To illustrate suppose that we have a model of 2 to the power 4 minus 1 and 4 third design model and it shows the that in which we considered the I a matrix or I value as the as the factors ABCD, suppose that after the data from the first 8 trials were observed the largest effects were basically ABCD.

So, we ignore the 3 factor interactions and A plus B, AB plus CD alias chains are ignored, the other 2 aliases can also be ignored, but clearly the AB or AB or both of 2 factors are large hence they cannot be ignored, to find out which interactions are important we could of course, run the on the alternative fraction which would require another 8 trials. Then all these 16 trials could be used to estimate the main effect and the 2 factor interaction and alternative would be to use a partial fold over involving 4 additional runs and we can complete the task accordingly.

So, again now, as we are considering the partial fold model so; obviously, the multiple linear regression would be larger in size and what is that please pay attention to here, it is possible to de alias AB and CD in fewer than 4 additional trails. Hence, we will assume the model as such initially what we had we had y is equal to because it was there was no fold. So, it was y is equal to beta naught plus beta 1, x 1 plus beta 2 x 2 plus beta 3 x 3 plus epsilon.

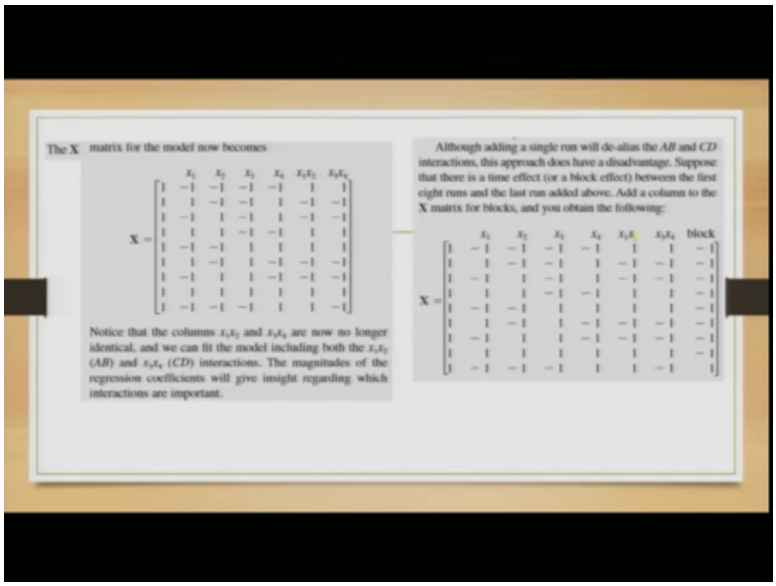
Now, the model is changing, what is it please pay attention it will be y is equal beta naught plus beta 1 what beta 1 into x 1 plus beta 2 x 2 plus beta 3 x 3 this remains as it is. Now, we are basically considering those extra factors as it should be as per the fractional factorial modules considering the partial folds, the fifth term would be x 4 into beta 4 plus beta 1 2 where you are combining the effect of A and B which is be x 1 into x 2 and the and th last term would be apart from the error term would be beta 3 into beta 3 4 which will be x 3 into x 4 plus epsilon.

Where the the factors x 1, x 2, x 3, x 4 are the coded variable represents ABCD and the combination of x one x 2 or x 3 x 4 with the combination of AB and CD. Using the design the x x matrix now be given by n remains the same, but the factors which is n cross m k plus one or n cross k plus 2 whatever it is would now depend on the factors.

Which we have consider, what are the factor it is x 1, x 2, x 3, x 4, x 1 and x 2 which is the fifth one and x 3 x 4 which is a sixth one. So; obviously, it would be n cross 7 because the 7 may, seventh means which is the first one would be beta naught. So, there would be beta 1, beta 2, beta 3, beta 4, beta 1 2 beta 3 4, where we have written the variables above the columns to facilitate better understanding noticed there is that x 1 x 2 column is identical to x 3 x 4 column ah, because AB or alias CD are coded as x 1 x 2 and x 3 x 4 respectively implying a linear dependence model. Therefore, we cannot estimate both x 1, x 2 and x 3, x 4 in the model, I was suppose that we add additional run and basically consider the x 1 x 2 as different levels.

We can basically have alternate fractions to the original 8 model, 8 runs and basically be able to estimate beta not beta 1, beta 2, beta 3, beta 4, beta 1 2 and beta 3 4.

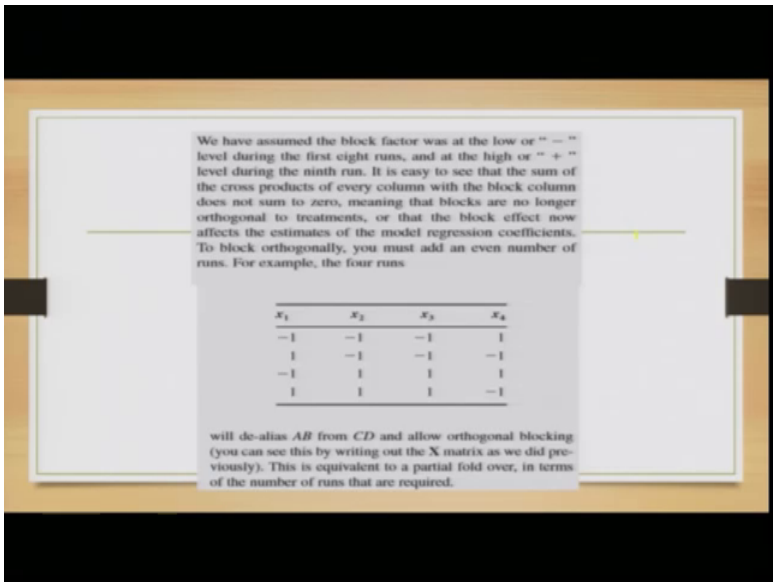
(Refer Slide Time: 18:58)



Once you have the other matrix it is again of size n cross, I am repeating the word n in order to me basically denote the number of observations, n cross 7. Notice that the column x 1, x 2, x 3 and x 4 are no longer, are now no longer identical and you can fit the model increasing and find to try to find out the estimates of beta 1 2 hat and beta 3 4 hat 3 4 basically this is a suffixes, the magnitude of the regression coefficients will give inside regarding which interactions an important.

So, once we do, do that we can basically do the same methodology we are trying to basically find out the beta hats, try to find out the errors, try to fit in the model to find out the average of the errors, try to fit and find out the distribution of the errors and everything and would basically remain the same. And; obviously, we will come into the main table which is for the ANOVA over where the reading on the variables of the attributes, I am using the attributes because it can be attributes and variables also depending on what type of model you are considering any of the sum of the squares you have the degrees of freedom you have the f values, the mean square value and values the f value and basically you have the level of significance.

(Refer Slide Time: 20:09)

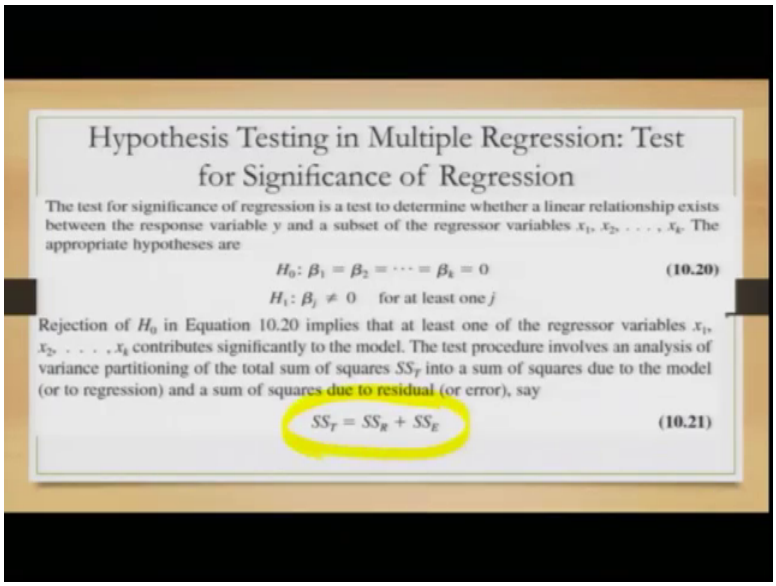


We have assume the block factors were at the level of minus in the first 8 runs then at the high level of during the ninth run it is easy to see that the sum of the cross product of every column with the block columns does not sum to 0, meaning the block are longer orthogonal to treatment or that the block effect.

Now, effects the estimate the model regression coefficients, to block orthogonality we must add an even number of runs for example, we have extra 4 runs where it will give us because there are what, there are x 1, x 2 and x 3, x 4 combined which is AB and CD. So, hence we basically add up more number of combinations for that, such that we can basically do the calculations and try to basically differentiate the effects in much better fashion.

So, with the analysis for AB which is x 1, x 2 and CD which is x 3, x 4 and for allow orthogonal blocking this is equivalent to a partial fold over in terms of the number of runs and that are required in order to estimate.

(Refer Slide Time: 21:08)



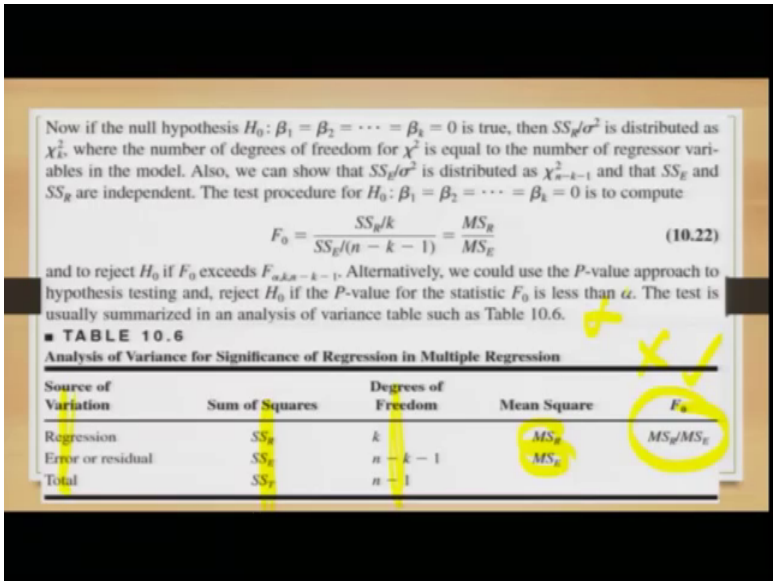
Now, you want to do that hypothesis testing, again we are going to come back to the ANOVA table. So, the test of significance for regression is a test to determine whether the linear relationship exists between the response variables y and a subset of the regressor variables x 1, x 2, x till x k. Now, the actual hypothesis would be stated which is in h naught case we will consider beta 1 to beta k all to be 0 and in h 1 which is the alternative 1 one cases we will consider that that for at least one of these j, j is 1 to k they are not 0.

So, in this case 0 means that the effects of them are 0; that means, we need you to the partial differentiation the rate of change of y with respect to beta 1, x 1, x 2, x 3 whatever is 0; that means, they are not being affected. So, rejection of h naught in equation in this equation implies that are at least one of the regression variables x 1 to x 3 contribute significance so; that means, if you are rejecting h naught been basically b tech taking cognizance of the fact that h a is true.

So, if h a is true; that means, one of the betas are definitely not equal to 0, which means that there is a linear relationship between y and that particular x. The test procedure involves an analysis of variance exactly as ANOVA table where we find out the total number of squares into the sum of the squares due to the model and due to the sum of the errors total sum is equal to sum of the variables plus sum of the errors. So, this is the, this

is the main thing which you have been discussing and from there we will decide we do the ANOVA table.

(Refer Slide Time: 22:48)



Now, if the full a null hypothesis which is H naught is 2 then SS R which is for the regression models divided by sigma squared is distributed to chi square, because chi square is basically distribution will consider with respect of sigma square and the standard error distribution, which we are considering where the number of degrees of freedom of chi square is equal to the number of degrees variables in the model.

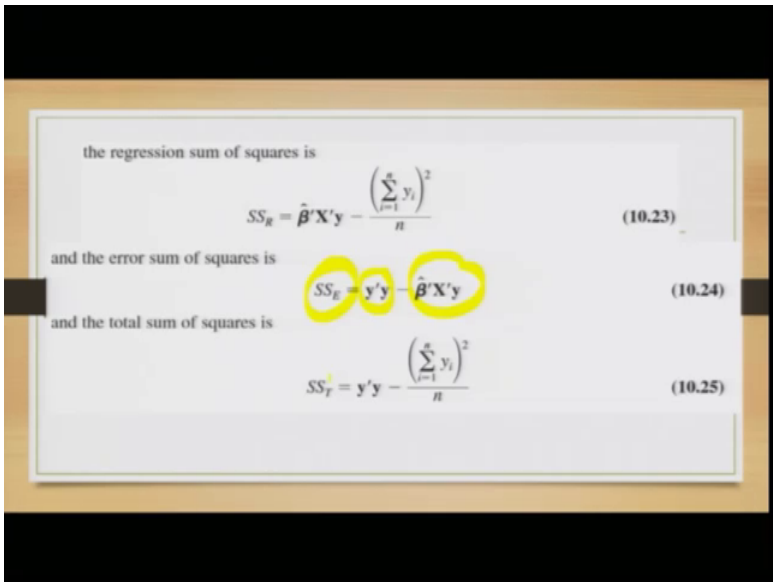
Also we know that the sum of the squares of the errors by sigma square is distributed, again chi square with the certain change in the degrees of freedom and that sum of the squares of the errors and sum of the squares of the total effects are independent. That is what is very important to note down, considering the fact that we are considering normal days, normal distribution to be true for all the x s errors and y, this is very important to be noted by all of you for basically listening to this lecture and who are taking this course.

The test procedure would be basically for H naught would be to find out if F statistics is a sum of the squares of the product of the case that by k, with k is the degree of freedom depending on the number of variables divided by sum of the square of errors is divided by it is degree of freedom. That will give me the mean square of r by divided by mean square of the errors and we will reject if it H naught, F naught under h naught exceeds

that value F alpha, alpha is the level of significance which you have alternatively we could use the p values to approach to test the hypothesis.

And reject F naught with p value of the statistics F naught is less than or if it is greater than alpha we accept that alpha can be 1 percent, can be 5 percent, can be 10 percent so on and so forth. The test statistic is usually summarized again in same ANOVA table. So, what we have is this here the regression errors in the residual are given in the first column, sum of the squares is given, degrees of freedom is given divide SS R by k is given, then the mean squared errors SS e by n minus k minus 1 will give me the mean squared errors based on that you can find out the SS F statistics and then we can comment with whether it is right or wrong depending on the alpha value which is the level of significance.

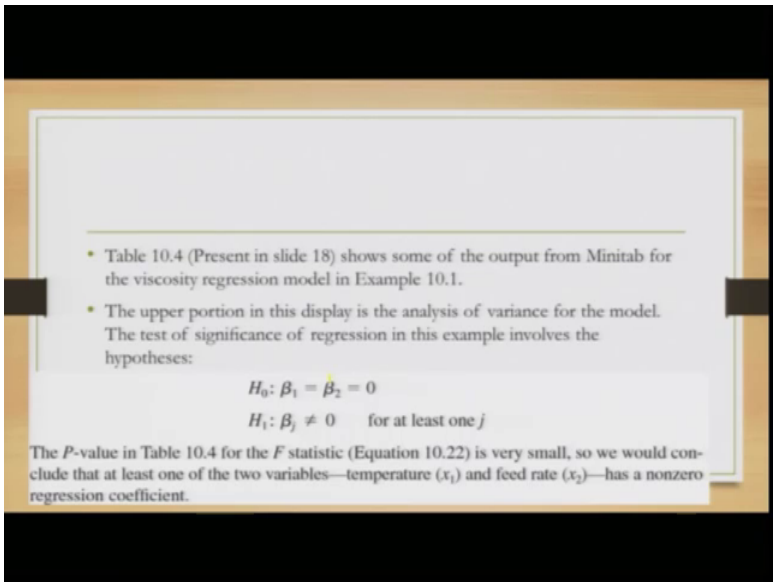
(Refer Slide Time: 25:08)



The regressions on the squares by the formula is given by beta hat transpose. So, beta hat what basically of size k plus 1 plus 1. So, it transpose would be just the opposite, multiplied by x transpose y minus the overall effect of y s, y s are the actual value divided by this.

So, this is the basically the best fit line which you have for the actual value and the sum of the squares would be given by, this is the total error minus this would give me the error sum of the square of the error. So, total value would be given by addition of them and you can find out the calculations calculate the values accordingly.

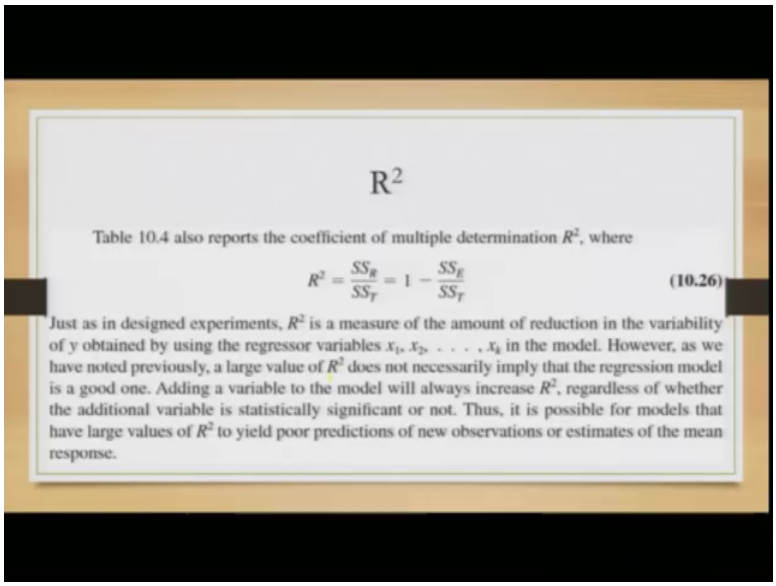
(Refer Slide Time: 25:53)



So, in the in the this following slide it will shows the sum of the output mini tab for the viscosity regression model for the upper portion of the displays is the analysis and variance for the model.

The test of significance of the regression in this example involves the hypothesis, that beta 1 is equal to beta 2 is given 0 and either and in the alternative one would be at least one of them is not 0, the p value of the table for the f statistic very small. So, you would conclude that at least one of the variables which is temperature or feed rate have a non 0 value hence the linearity relationship holds.

(Refer Slide Time: 26:31)

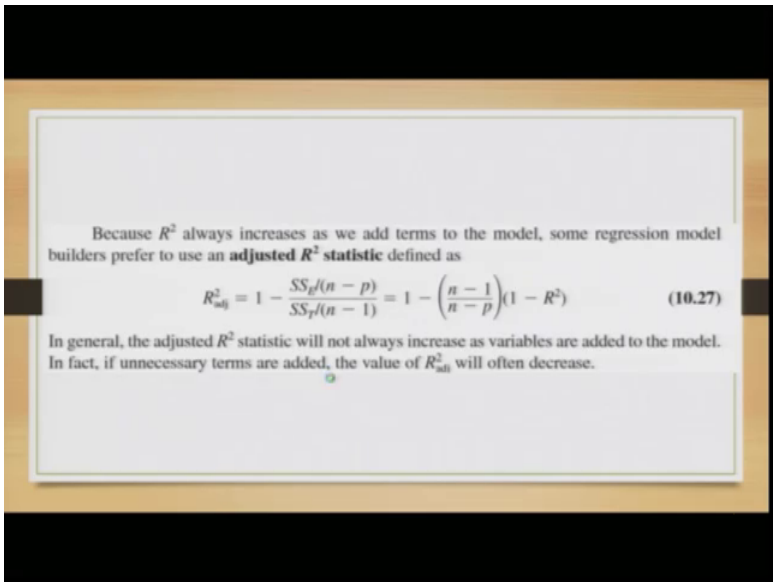


The table also gives a R square value, which is that adjusted R square and adjusted R square, R square would be the sum of the squares which are we are able to predict from the variables divided by sum of the total of the total errors. So, that would also be equal to 1 minus, because the ratio is always one for the total was the total minus the sum of the square by the error by the total.

Just in design experiment R square is major amount of deduction the variability of y obtained by using the regression variables x 1, x 2, x 3 to x k; however, as we have noted previously a large value R square does not necessary imply the regression model is a good one. This is interesting and listen to it adding a variable in the model will always increase r square; obviously, because we are adding more and more variables we are able to think that we are able to think more and more. Regardless of whether the additional variable in statistically significant or not, thus is it is possible for models that have large R values to yield poor prediction of the new observation or estimates of the mean values as we proceed for the n th plan predictions to n th predictions and so on and so forth.

So, basically we should take the minimum number of x s try to predict the maximum, basically have a value of R square where it will be able to give us the predicted errors as least as possible for the for the n th plus 1 n th plus 2, n th plus 3 data.

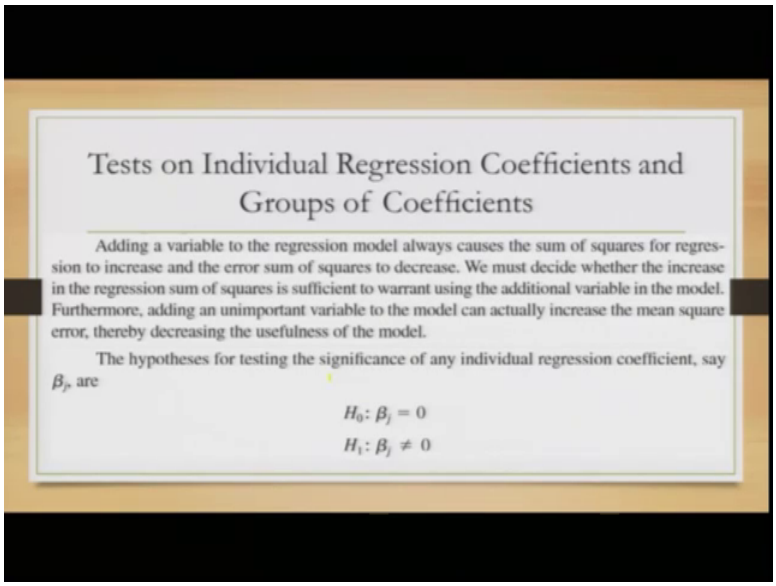
(Refer Slide Time: 28:00)



Because R square always increases as we had terms in the model some regression models builders prefer to use adjusted R square which I mentioned is basically by dividing by the degrees of freedom. So and adjusted r square would be given 1 minus so, initially it was sum of the square of the errors divided by same squares of the total.

Now, it will basically be n minus 1 by n minus p p is basically k plus 1 depending of beta not is there or not multiplied by R 1 minus R square in general adjusted R square statically R not always increases variables are adjusted to the model. In fact, if necessary terms are added, the value of R square adjusted would often decrease and basically give us an actual value of the prediction level of the model.

(Refer Slide Time: 28:41)



So, now we will basically start of the test and individual regression coefficient and the group of coefficients. So, adding a variable regression model always causes the sum of squares for regression module. So, increase in the errors of the sum of of of the squares to decrease.

So, obviously, we want as x s are increasing R square would basically be give us the maximum output, but; obviously, the question remains that to what level are we able to predict for the n th plus 1 n th plus 2 data points that is point number 1. Whether the errors as per the assumptions have normally distributed with the 0 mean and certain variance whether the non on normality assumption holds and the covariance of the beta which you want to find out are really small or not.

Because if they are very large; obviously, apart from un biasedness the consistent property would not hold, we must decide whether the increase in the regression sum of the squares is sufficient to warrant using additional variable in the model. So, we will try to basically find out the hypothesis that other than testing the significance of any of them collecting as a group, we will try to take them as an individual and try to predict as maximum as possible.

With this I will end the 38 lecture and try to continue with this combinations of the hypothesis testing such that we take individually the beta and try to predict with H naught is true or H a is true and basically pass judgement accordingly.

Thank you very much and have a nice day.