Total Quality Management - II

Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 37
Fitting Regression Models – II

Welcome back my dear friends, a very good morning, good afternoon, good evening to all of you and this is the TQM-II lecture under the NPTEL MOOC series and we are in the 37th lecture; that means, we are in the second day or the second class for the last week. And I am Raghunandan Sengupta from IME department IIT Kanpur.

So, if you remember you in the last class we did not cover much of the slides, but they were lot of discussions which you are basically given and hope it is clear to all of you. So, we were discussing in the multiple linear regression and did mention in very few words why multiple linear regression means the linear part of the betas and not the non-linear part of the betas.

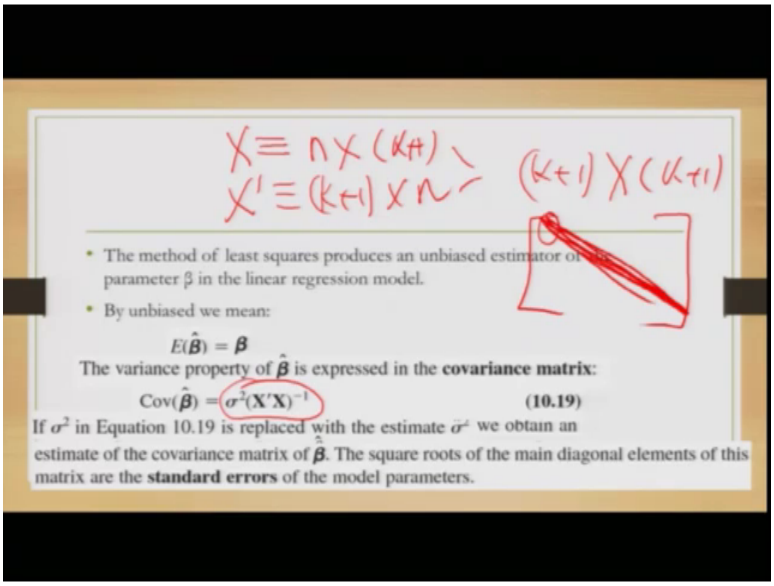
Because we are going to partially differentiate the sum of the square of the errors with respect to the unknown parameters with the betas, put them to 0, find out using the normalized linear equations, find out the betas, use this betas to estimate the values of y for the nth plus, nth plus 2, nth plus 3, nth plus 4 reading, then find out the difference of the using those errors; sorry using those errors, errors means the difference between the y actual value and the predictive value and where using the values of beta hats; whatever there is beta naught, beta 1 beta 2 till beta k.

So, those hats we utilize to predict the values of a y which have the y hat values find out the difference between the actual y and the y hats, predicted values of y hats; find out the errors, sum of the errors. In the long run should be 0 because the expected value of the errors is 0, the variance should be 1 or sigma square as the case may be and then we proceed.

And then we find found out and we did discuss that how we can find out the betas using the simple concept of matrix notation? How we can find out the sum of the squares? How we can find out the sigma square? And at the fag end of the class of the 36th lecture I did not mention that de concept of degrees of freedom; that means, you are losing set of information's. So, if there are 2 betas you will you will basically use those set of observations to find out beta naught, beta 1. If you have 3 betas you will basically use them for 3 times to find out beta naught, beta 1, beta 2.

So, hence you are losing the degrees of freedom. Hence the sum of the square whatever the formulas which you use main part of the discussion was basically you are going to divide by n minus p; p was basically k plus 1, where k plus 1 is the number of betas you are going to estimate. So, you are losing information's. Hence the degrees of freedom would be n minus p or n minus in the bracket k plus 1 and n is the set of observations.

(Refer Slide Time: 03:00)



The method of the least squares produces an unbiased estimator. I did mention the unbiased estimate in the 36th lecture. So, produce an unbiased estimate of the parameter beta in the linear regressions. So, what we mean by unbiased estimation is the expected values of the beta hats in the long run should be exactly equal to the values of betas which is the example 2, very simple examples which I gave was keep tossing a coin; number of heads for each such set of tosses which you have.

If you have say for example, 200 tosses find out the and the chances of heads in the first 200, then chances of head on the second 200, third 200 so and so forth, add them up divide the number of 200 such tosses you have done, find out the average of the average. It should be exactly be equal to 0.5. So, each actually means the unbiasedness.

So, by unbiasedness we mean that the expected value beta hat should be equal to beta. Now the variance property of beta is expressed in the covariance matrix. So, we if you basically have the so, betas are what? In the in the vector notation the first I would beta naught, beta 1, beta 2 till the last term is beta k. So obviously, they would be covariance between them. So, if you want to find the covariances and use simple algebra matrix notation, the covariance structure would be like this, your sigma square which you already found out now here the term which we have.

So, if you have what was basically explain. So, if you know comes let us come to X basically X was n cross k plus 1 and X prime was k plus 1 cross n. So, X prime cross X if you multiply them you will basically have k plus 1 cross k plus 1, transpose of inverse of that will also be k plus 1 into k plus 1. So, when you find multiply sigma square sigma square is a scalar. So obviously, the actual matrix of the covariance of beta is basically matrix of k plus 1 k plus 1. So, of which would be right because as we have been discussing.

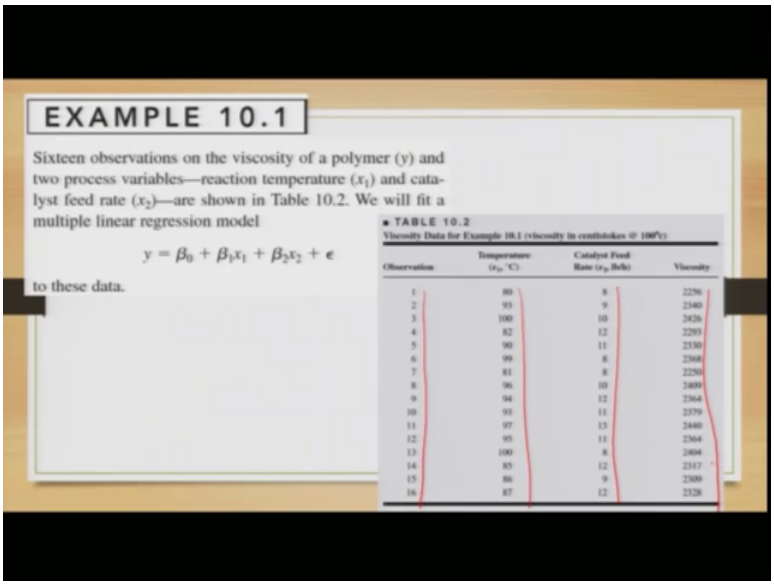
So, what we have the covariance of betas, the primary or the or the principal diagonal would basically the covariance of beta naught with beta naught, itself which is the first element here. The second element cross along the principal diagonal would be the covariance of beta 1 with respect to beta 1; which is basically the variance of beta 1. The third element along the principal diagonal would be the covariance of beta 2 with respect to beta 2; which is the variance of beta 2 and the last element which come along the along the principal diagonal which is the k plus 1 comma k plus 1 cell number would be the covariance of kth beta with respect to the of itself which is the variance of beta k and off the diagonal element with the covariance's.

So, the so the element 1 comma 2 and 2 comma 1 would be mirror image would to would with which would be the covariance of beta naught with respect to beta 1 in the element 1 comma 2 and in the element 2 comma one it would basically the could be the covariance of beta 1 with respect to beta naught, so obviously, there are mirror image So, this will be a give us a covariance's.

So, if it sigma square in the equations as we have just for discussed that in the end of the 35th lecture is replaced with the estimated of values of sigma square, we obtain an estimate of the covariance matrix of beta naught, beta hat. The square root of the mean diagonal elements of the matrix, as I just mentioned few seconds back, it is the standard errors of the model parameters or standard errors which basically; standard error is the word which we are using first time but it basically means the standard deviation of the sample.

So, when we are talking about the standard deviation we are talking about the population, when we are talking about the standard error which basically the standard deviation of the sample and we are talking of the variances; it will basically the variance of the population and we are talking about the variance with the word sample variance it basically would be the corresponding to the sample. So, the principal diagonal if is the covariance variance matrix and they you have found the square root. So, the principal diagonals with the standard errors of beta naught, beta 1, beta 2 till the last element and off the diagonal element are would basically with covariance's of the beta i and beta j; where i and j basically where changed from 0 to k.

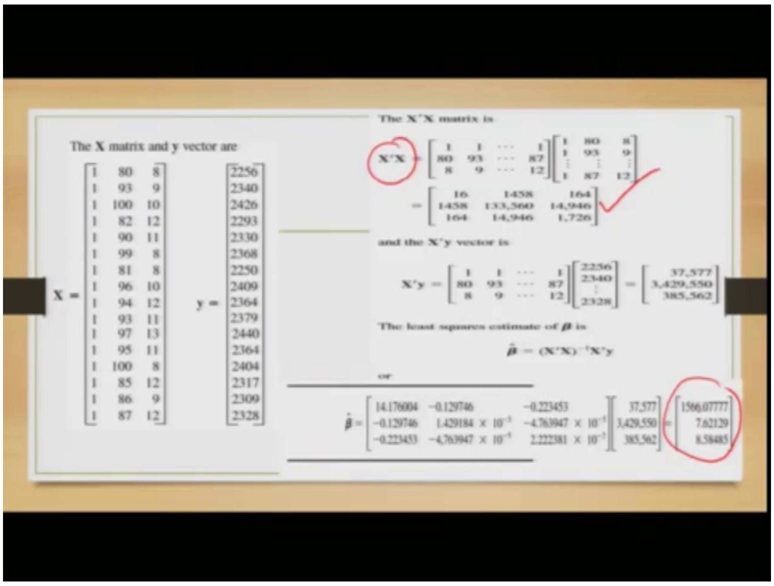
(Refer Slide Time: 07:44)



So, let us consider the example, 16 observations on the viscosity of the polymer. So, what is the y value? Is the viscosity of the polymer which you want to estimate of focused and 2 process variables are being utilized which is x1 is the temperature and x2 is the feed rate of the catalyst. So, these are the parameters based on which we will try to find out the viscosity of the polymer. The equation which we want to basically estimate, is the multiple linear regression of the form y is equal to beta naught plus beta 1 x1 plus beta 2 x2 plus epsilon; where why I am again repeating is the viscosity of the polymer; x1 is basically the reaction temperature and x2 is the feed rate of the catalyst.

So, the observations are given 1 to 16, the temperatures are given which is x1 in degree centigrade from starting from 80 to 87. The catalyst feed which is given in the rates of pound per hour is from 8 to 12. So, I am just reading the first value and the and the last value. So, there are 16 readings and the viscosity values which is the y values are given from values from 2256 to 2328. So, this is the viscosity data for the example and the viscosity is given in stocks. So, this is the in centistokes; which is stokes into 10 to the power minus 3 and when we have this so the matrix are, if you remember the metric x because there is beta naught.

(Refer Slide Time: 09:16)

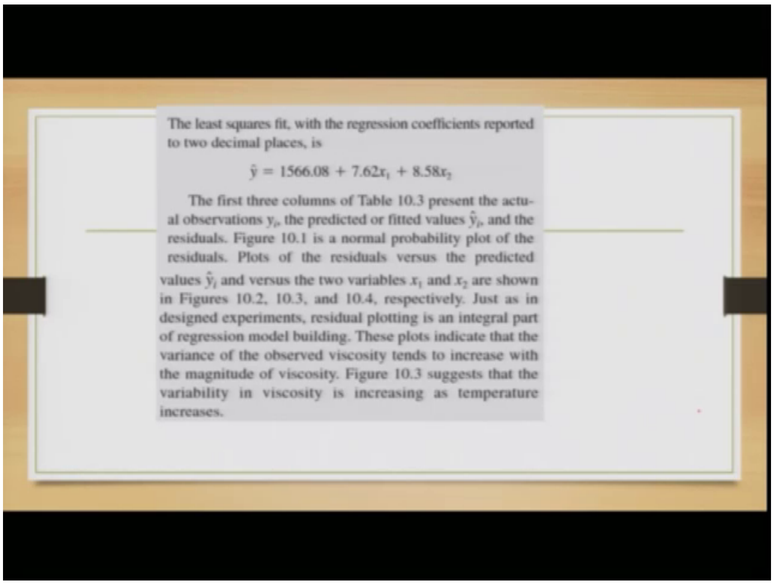


So, the first column would be the one ones. So, how many ones they would be 20 observations. So, on and the observations on the second column would be based basically corresponding to x1, observation in third column would be corresponding to the x2 and the y's are the centre stoke values of the viscosity starting from as I mentioned 2256 is the first value which is y1 till y20 which is 2328. So, once we have that we can find out very simply in the x matrix; then you can find out the x transpose matrix because it transpose would be rows would be columns, columns would be rows. Once we find then so, x cross x matrix is given and then we find out the x so, the if you check it, the size is 3 cross 3.

Why 3 cross 3? Because beta naught, beta 1, beta 2 and x transpose into y the value is given. So, again the size would be 3 cross 1 as we have been mentioning. So, once you find it, then put them in this equation. What was the equation? Equation was basically x into x prime x sorry x prime into x that multiplied find its transpose multiplied by x prime multiplied by y. In the sequence you will find out the 3 cross one values which is corresponding to beta hat.

So, that beta hat comes out to be this; which means beta naught is equal to I will not read the decimals. It would be 1566 is the value of beta naught, 7.6 let me bring the decimal values for beta 1. So, it would be 7.6 for beta 1; which means that for the change of the temperature the rate of change of the viscosity for that that polymer, in centistokes would be of the unit of 7.6. So, one unit change in the degree centigrade temperature would increase or decrease the value of the viscosity level by 7.6. Similarly, if you have the polymer a catalyst values, so there the this values are given which is basically beta 2 would be of d of a value of plus 8.5.

(Refer Slide Time: 11:47)



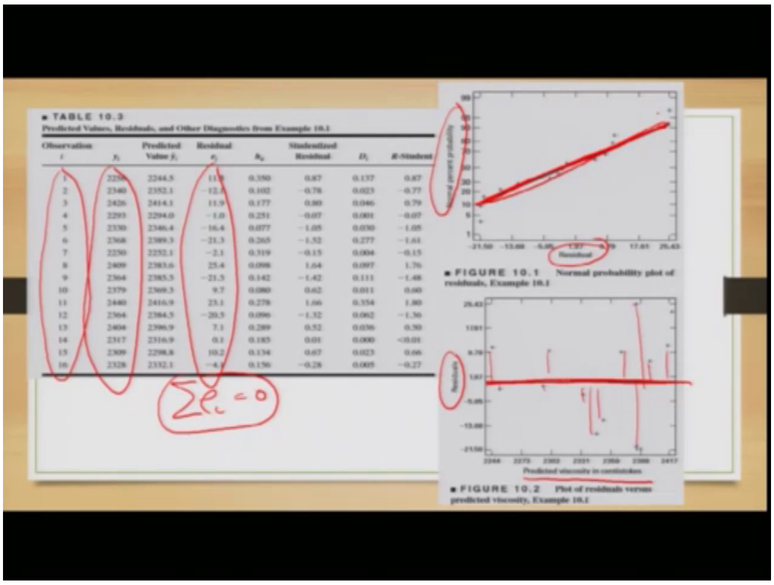
So, the least square fit with the regression coefficients as given here; would be y hat. Now this is y hat remember that it is not y actual value and also remember that the error terms would not come now until unless you find out the difference between y and y hat. So, we are need to find out the values of the predicted values. Why predicted values because we using the beta hat values. So, beta naught value which you have already found out is 1566.08. I am just reading till the second place of decimal, beta 1 is basically beta 1 hat is basically the value of 7.62; beta 2 hat is basically values of 8.58. So, beta naught multiplied by x1. So, what is x1 you will try to find out? It will be the 21st reading. So, beta naught plus beta 1 beta naught hat plus beta 1 hat multiplied by x1 which is the 21st value, then plus beta 2 hat multiplied by x2 which is the 21st value; that will give me the y hat value of the 21st value. So, y actually for the 21st value minus the y hat for the 21st value would give me the error for the 21st value.

Similarly, when I multiply beta naught plus beta 1 hat plus x1 for the 22nd value plus beta 2 hat into x2 for the 22nd value that will give me the y hat for the 22^{nd} value. So, y actually for the 22nd value minus y hat for the 22nd value would give me the error for the 22nd value. If I continue doing that I will find out the errors. If I find out the sum of the errors, it should technically be 0 because as the assumption we can find out the errors, the variance is a errors and do all the calculations as needed. So, the first three columns in the given the table represents the actual observations y y i's. So, predicted or fitted values of y hats are given and the residuals are also given; residuals mean the errors.

Now, if you plot the this errors. So, the normally plot of the residuals can be done. So obviously, they should be normal because the reason is that we have assume the x is are normal y is a normal. So, if there are normal the error term would also have a normal distribution with the 0 mean and a standard deviation which a fixed. Lots the residuals versus the predicted values of y hats and versus the 2 variables x1 and x2 are shown which is going to come to that. Just as in a design experiments residuals plotting is an integral part of the regression model building.

These plots indicate that the variance of the observed viscosity tends to increase with the magnitude of the viscosity and figures later on will suggests that the variability in the viscosity is also increases on the temperature increases. So, what I have just discussed I am going to now show it in a table or the other graph plot.

(Refer Slide Time: 14:37)



Ok my mistake if they were 16 readings, my mistake I just mention that 20 values. So, observations are basically 16 in number. These are the y values which have said. So, basically they start from 2256 to 2328. The predicted values we have found out. So, the first predicted value is 2244. 5, then the second values it 2352.1, the last value is 2332.1, the error residuals ah errors are found out.

So, technically when I mean that if it sum up the errors it should be should be 0. Then I find out the students is residuals if I find out, then I find out the R values and find out the degrees of freedom and I can do the calculations accordingly. So, the normality plot of the residuals. So, normality plot would be let me mention it I think I have mentioned it, but I will just repeat it.

So, what you do is that, this will take about 5 minutes. So, let me discuss. So, consider I am of the discussion and I am just stepping out of the discussion and coming out to how we do the q q plots of the quantile-quantile plots. Consider you have a distribution and for the time being consider it is normal distribution which we know. What we know is basically is mean and the and the variances and we know all the readings, so for the normal distribution and that is basically from a population; that means, there in financial of solutions, that is kept on one side.

On other side you have a unknown distribution which you want to fit and find out whether is it actually equal to the normal distribution. So, consider that distribution is get the observations given that is kept on the other side on the right hand I consider and the normal distribution a take I long them from the least to the highest and plotted along the x axis. So, these are all marked on the x axis's theoretically. Now I take the other distribution unknown distribution which is on the right hand side, rank them from the least to the highest. One when I mean the rank them, when I rank them I also have the probabilities obviously. Now what I to do is that I find out the quantile values. So, 10 percent value, 20 percent value, 30 percent value so and so forth.

Now, also an I basically plot them along the y axis. Now draw a 45 degrees line in the graph. Now if the normal distribution quantile, so each step I am going to take for the normal distribution, if the same properties being covered in the in the unknown distribution also; so obviously, those one to one correspondence would basically mark the plots along the 45 degrees line theoretically; which means the distribution based on which I am trying to plot is quantile-quantile plot and trying to compare that unknown distribution with the normal distribution as they match. Hence we can say the unknown distribution is also normal distribution.

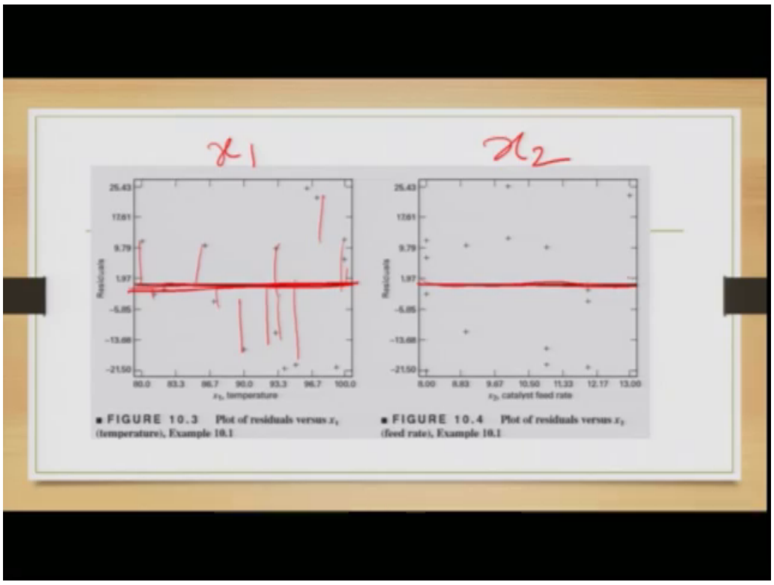
So that that case, now in the case if the values are above this 45 degrees line or below this 45 degrees line would give me the idea, whether this is cued on to the left or the right with respect to the normal distribution. So, this known distribution or normal distribution which I said that we should be kept on the left can be in place of normal distribution, can be exponential distribution. So, we can compare a known exponential distribution with an unknown distribution and if it matches, then we can say the unknown distribution is normal. So, they can be done we can do the q q plots or the quantile-quantile plot for the exponential distribution, the gamma distribution, the Poisson distribution and so and so forth.

So, in this case we assume and rightly so, that the residuals have a normal distribution. So, if you plot them this is a 45 degrees line which we are talking about. So, this is the residuals and this is the normality plot. So, this can be done on the on the x axis and the y axis whatever. I did mentioned we do it on the x axis, but it can be done on the y axis and the residuals plots are there. So, if we find out this point which are there they are almost falling; obviously, they are not exact. They are along the 45 degrees line. So, we can safely assume that the assumptions and normality wholes for excess as well as for the errors.

Now, if you plot the residual versus the predicted viscosity, so obviously that will give me the errors. Now why I am plotting that, because if you remember I have said sum of the error should be 0 in the long run. Because as the expected value of the errors. So, here I have the predictor viscosity in centistokes predicted y hat and this is the residuals. So, basically I have the residuals being plotted. So, this is plus minus plus minus, if I add up the sum of the errors.

Obviously, they in the long run should be 0 because the expected values 0. These are the way of trying to basically double check the overall concept which I am doing. Two important things which I did mention time and again; the expected value of the error should be 0, not may not be exactly 0, but it should be as close as possible to 0. So, if I find out the error terms of this plotted would be almost equal to 0 value, where would the 0 value would be somewhere here. So, this is what it should be and the error should be basically normal distributed. So, if I basically do the q q plot it should be along the 45 degrees line.

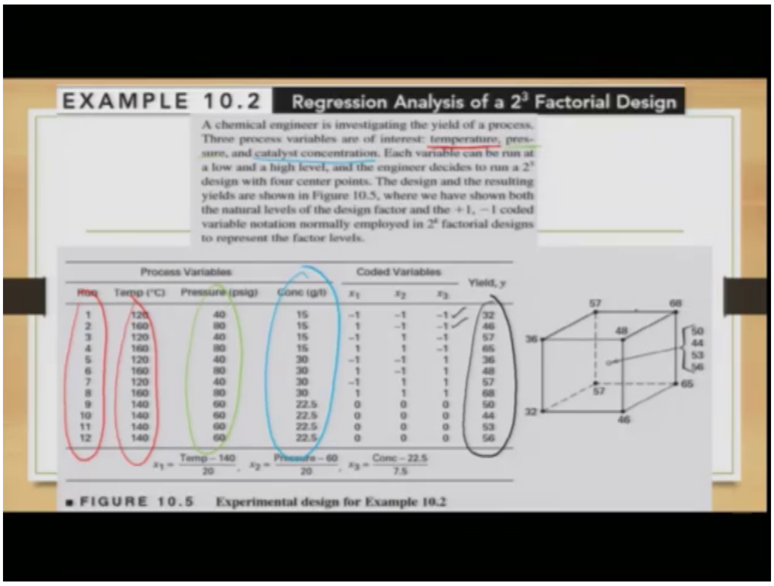
(Refer Slide Time: 20:19)



Now, I will do the plot of the residual versus x1; which is the temperature. Then I do the plot of the residual with respect to feed rate; which is pound per hour. So, here also you find out. So, I this is the average value which I am plotting. These are the deviations. So, in the long run should be 0 as it is, it should be also 0.

So, this is with respect to x1, this is with respect to x2 and you can do it for y hat and so on so fort. I am plotting the errors on the y axis y axis, in the x axis I am taking the predicted values of y once the x1 values, once the x2 values and so on so forth. So, these are basically double checking; the assumptions are which are expected value of the errors and the concept of normality plot.

(Refer Slide Time: 21:23)



Now it is basically the Regression analysis of 2 to the power 3 Factorial Design. So, there are 3 factors of the level each being of level 2 plus minus 0112 whatever. A chemically engineering investigating the yields of a process. Three process variables are of interest. That is why it is 2 to the power 3 those, that 3 basically means x1, x2, x3; which are temperature, pressure and catalyst concentration.

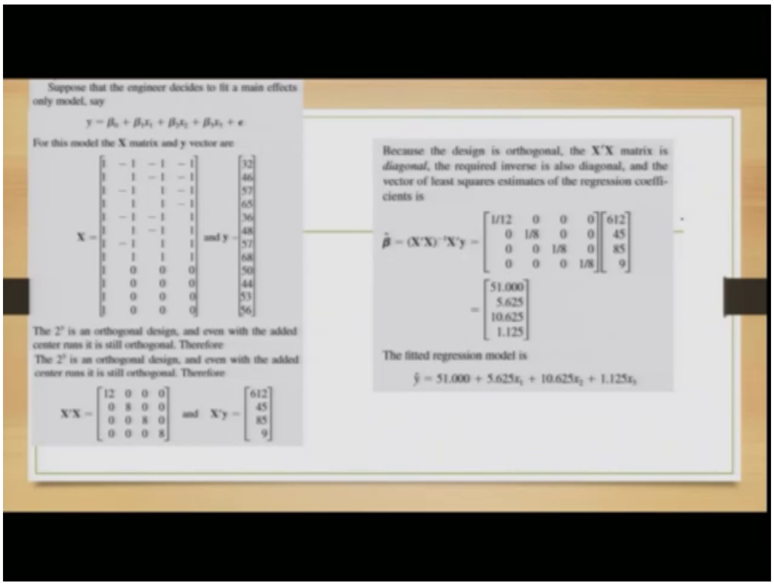
Each variable can run at a high and a low value. That means what we say low and high 0, 1, 1, 2 whatever. So, if it was basically 3 to the power 3; so that means, 3 catalyst values at 3 different levels. Each variable can be run at low and high level as I mentioned and the engineer decides to run a 2 to the power 3 design with 4 center points.

The design and the resulting yields are shown in the figure 10.65 which are the variables and the values; where you have shown both the natural levels of the design factor and the plus 1 minus 1 coded variable notation, normally implied to root as in the 2 to the power k; factorial fractional design concepts. So, the runs are given, the temperature are given. So, temperature is basically the one of the variables. So, if I go to the so, the pressure is given. So, the pressure in pound to per squares inch, if I go to the catalyst concentration catalyst concentration is given.

So, now this variables are coded. So, the codings are done as so, plus minus of 0 1, 0 1, 0 2, 0 2 and then 01 and 10. So, the variables are given. So, this is at a low lower level x1, low level x2, low level x3. So, this is the first one. Let me use a black color. So, it will be easier for me to; so, this one is basically are at low level, this one is that first one is high level, the second two are in low levels and so and so, forth and then we have the yields. So, yields is basically the y variable we are trying to find out.

So, the temperature is the average value is basically given; the pressure average value is given, the concentration is given and we basically draw the relationship between the 3 variables in a 3 dimensional figures. So, where along the x y x axis, y axis and z axis; you will basically plot x1, x2 and x3. So, now the question is of the form y is equal to beta naught plus beta 1 x1 plus beta 2 x2 plus beta 3 x3 because there are 3 variables corresponding to x1, x2, x3 and beta naught for the case where you want to find out that the concept that if you remember y is equal to m x plus c. So, that concept where trying to fit plus epsilon.

(Refer Slide Time: 24:57)



So, now, if you are, if I find out the size of x so; obviously, the first column would be corresponding to the 1; all the ones which is corresponding to beta naught. The second column would basically be corresponding to all axis x1; the third column with respect to x2; fourth column with respect to x3 and y values we are already noted down. So, if you basically have x which would be n cross 4; n is the number of reading.

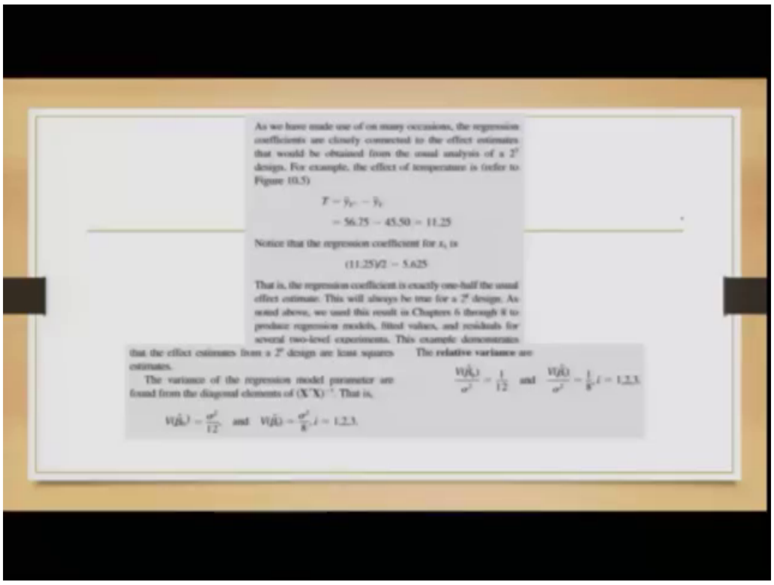
So, that can change and obviously, n has to be more than 4. Why 4 because beta 1, beta naught, beta 1, beta 2 and beta 3. Similarly y would be the size of n cross 1. So, the 2 to the power 3 is an orthogonal design and even with the adders central runs it is still orthogonal. Therefore, the 2 to the power 3 can be solved using the simple concept of regression model. We find out x transpose x which is basically of size 4 cross 4.

Then we find out x prime y which will be the size of 4 cross 1 and once we find out the values of beta naught's are found; beta hats are found out. Because the design is orthogonal, the x transfer x matrix is a diagonal matrix. The require inverse is also diagonal and the vector of lead squares estimates the regression coefficient and we find out the regression coefficient as beta naught as 51.00, beta 2 is basically 5.62. I am only reading till the second place of decimal; beta 2 is basically 10 of these are hat sorry my mistake.

So, be I will again repeat; beta naught is basically 51.00, beta 2 hat is basically 5.62, beta 3 hat is 10.62 and beta 3 is basically be hat is basically 1.12. So, this value and obviously, the errors would not be there. We fit find out the nth plus 1 reading of y; that is y hat, suffix n plus 1, find out the errors, which is errors have n plus 1; find out the errors for n plus 2; n plus 3 so on and so forth add them up, check with there is equal to 0.

And you can do the double checking on the normality plot, the q q plot, the residual plots, check out the whether the q q plots sum matching with the 45 degrees line, check whether the average of the errors are 0 because you are going to plot out the residuals with respect to ones with respect x1, ones and with respect to x2, ones with respect to x3 and then you are trying to plot the q q plots of the values of the predicted values or y hat values or the error terms with respect to the values of a of a theoretical normal distribution.

(Refer Slide Time: 27:28)



As we are made use of on many occasions, the regression coefficients are closely connected to the effect estimates that would basically be closely connected to the effect estimates that would be obtained from the user analysis of 2 to the power 3 analysis design. For example, the effect of temperature is to refer and we find out the temperature values; it can be either plus or minus because that was on the variables. So, for all these cases we can find the difference of the temperatures. So, notice the regression coefficient of x1 which is respect to the temperature is given by 5.625. So, because you are you have made them orthogonal. That is the regression coefficient exactly 1 half the residual effect.

This will always be 2 for the 2 to the power k design because there are 2 levels of for any factors, there 2 levels. As noted above we use this result to produce the regression models and fit in the values and residuals. This example demonstrates the effect estimate for the 2 to the power k design on the least square values. So, from that we can find out the errors of respect with respect to be the variance of beta naught hat, beta 1 hat, beta 2 hat, beta 3 hat and they come out to be as given the values. So, the beta naught hat have seen the error is sigma square by 12. So obviously, sigma square is the actual variable. If you estimated we have to replace by estimated value. While the beta 1 hat estimate variance, beta 2 hat variance, beta 3 hat variance all come out to be sigma square by 8.

Now, this variances when you find out the square root; this is the standard error if you remember I did mention that. The relative variance can be found out and we can basically solve it and basically have a nice idea how the regression models can be utilize for this type of ANOVA models. Now obviously, we will again at the end of the day you will have basically the number of readings, the sum of the squares, degrees of freedom, the f values and whether they are matching with the p p to what level degrees of freedom, what levels of confidential you are able to predict.

So, that the ultimate table is basically the ANOVA table, that will be can utilize here also. So, with this I will end the 37th lecture and continue the discussion of the last 3 lectures in with respect to the more further models of regression ANOVA models and few remaining topics which are there in this TQM 2 course.

Thank you very much and have a nice day.