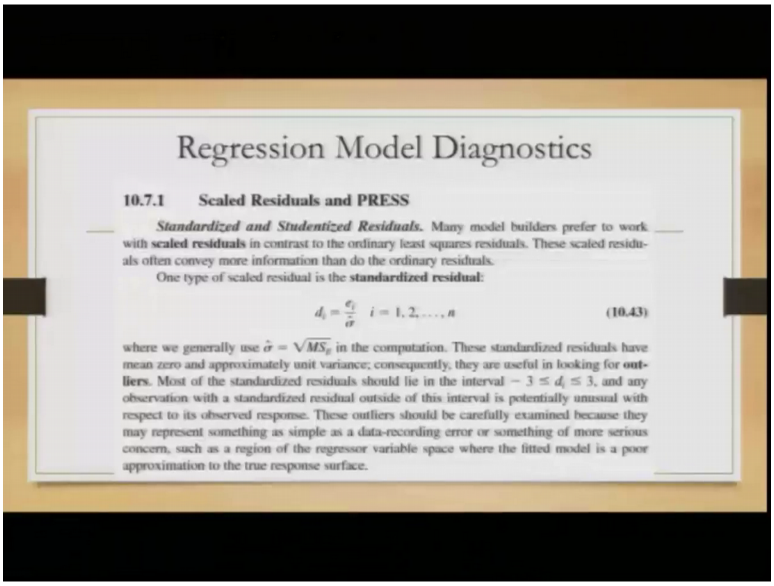
Total Quality Management - II
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 40
Fitting Regression Models – V

Welcome back my dear friends. A very good morning, good afternoon, good evening to all of you; this is the last lecture for the TQM-2 class under the NPTEL MOOC series and I am Raghunandan Sengupta from the IME department IIT Kanpur and I am sure that even though I should cover a few things, it has been a long journey they have been lot of coverage of concepts and I will try to wrap it up in the last few minutes, but I will discuss few pending issues and then as we will go into the part for the closing this TQM-2.

Now, if you remember that we were discussing that as addition and deletion of variables; what is the change of the sum of the squares, whether you are going to accept or reject h naught and the that would be under either the t value as a tt statistics or the f value of the f distribution and we will consider that accordingly.

(Refer Slide Time: 01:21)



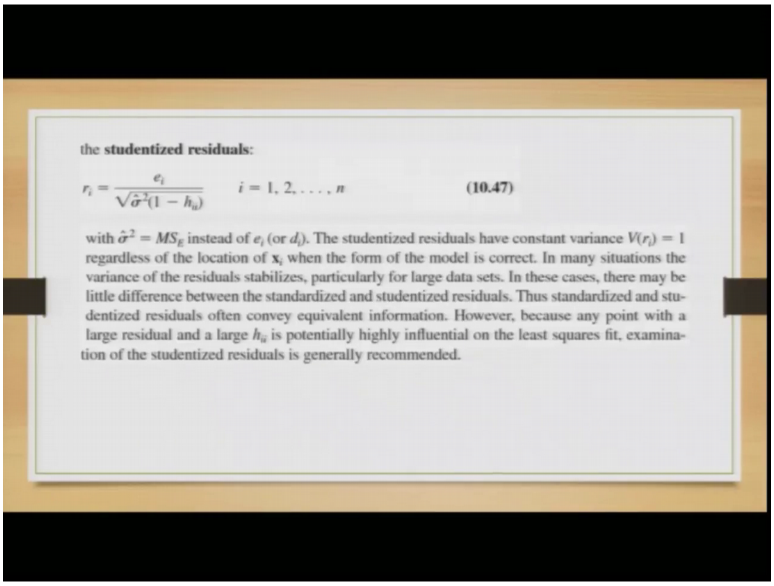
Now we will consider the standardized and student students studentized residuals; many model builders prefer to work with scaled residuals. So, they are being scaled or normalized to a value like you will be giving a marks for the course on a absolute scale or a relative scale.

These scaled residuals often convey more information than do the ordinary residuals. So, the scaled residuals of the errors divided by the divided by the sigma square and sigma square is basically hat, sigma square hat or always basically the mean square, square of the mean squares of the errors. These standard results have mean 0 and approximately you need variance because you are considering normality so obviously, this would be a consequence.

They are useful in looking for outliers most of the standard residuals should write in the intervals of plus minus 3; plus minus three considering that you are you are in a position to consider about 99.97 percentage of the on the variability that is plus minus 3 sigma and any observations with the standardized residuals outside the interval it potentially unusual and there are outliers they can be rejected or action can be taken accordingly.

These outliers should be carefully examined, because they mainly represent something as simple as a data recording errors or some something of a more serious concern such as the region for the degrees of variable space when the fitted model is poor in its prediction like, if I mean if you remember in the last part of the 39th lecture I did mention the in sample and out sample.

(Refer Slide Time: 02:59)



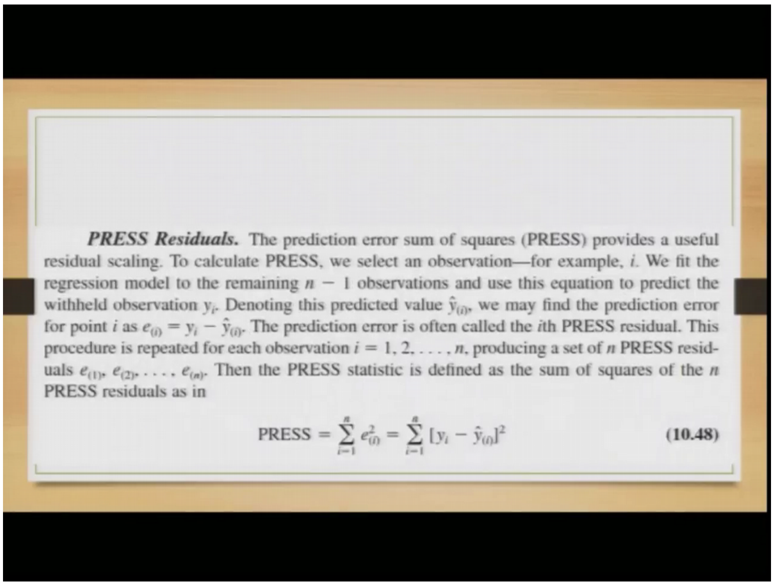
So, the student studentized residual would basically be considered using the end ratio of the errors of the i ith element divided by the square root of sigma square into one minus h i i.

So, what is that I will come to that later? So, this is the student ah standardized version of the the values which you are taking for the errors considering the in the covariances of beta a, beta j but they are now standardized. The students studentized residuals have constant variance; which is 1 regardless of the location where it is? What is the mean value of x s.

In many situation the variance of the residuals of stabilizes depending on more and more data you have so obviously, for less number of data the variability or considering the concept of consistency would be very high as you take more and more data always the would be they would be cluttering in around the mean more and more. In this cases there may be little difference between the standardized and studentized residuals.

The standardized and studentized residuals often convey equivalent information; however, because any point with a large residuals and large values of h h a i is potentially highly influential on the least squares for it; examination of the studentized residuals are generally recommended in order to pass much more scientific comments about the observations.

(Refer Slide Time: 04:25)



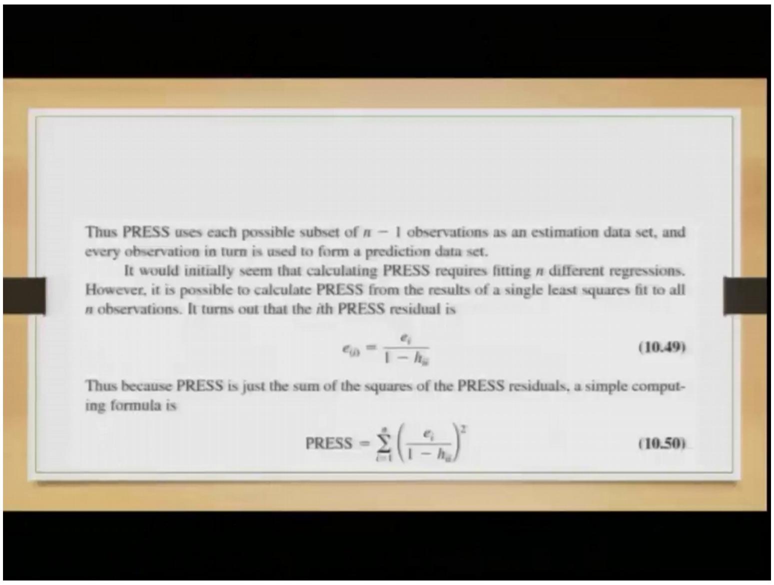
PRESS Residuals. The prediction error sum of squares (PRESS) provides a useful residual scaling. To calculate PRESS, we select an observation—for example, i. We fit the regression model to the remaining $n - 1$ observations and use this equation to predict the withheld observation y_i. Denoting this predicted value $\hat{y}_{(i)}$, we may find the prediction error for point i as $e_{(i)} = y_i - \hat{y}_{(i)}$. The prediction error is often called the ith PRESS residual. This procedure is repeated for each observation $i = 1, 2, \ldots, n$, producing a set of n PRESS residuals $e_{(1)}, e_{(2)}, \ldots, e_{(n)}$. Then the PRESS statistic is defined as the sum of squares of the n PRESS residuals as in

$$\text{PRESS} = \sum_{i=1}^{n} e_{(i)}^2 = \sum_{i=1}^{n} [y_i - \hat{y}_{(i)}]^2 \qquad (10.48)$$

The prediction errors sum of the squares so you want to predict this the sum of the squares errors its provides a useful residual scaling. So, you are basically you want to scale down the positive and negative fluctuations and you want to predict it; to calculate this value we of select an observation for any random observation ith one; we fit the regression model for the two the remaining n minus 1 observations and use this equation to predict and with that value predict that actual value of y i. I can many thing 1 2 3 4 5 6 till n 1 and predicted within the corresponding had value.

Denoting this predicted value as i y hat we may find out the error; the prediction it is call called all also called the ith error for the value of the prediction sum of the squares errors. This procedure is repeated for each end of the observation 1 2 3 4 till n and obviously, we will have the errors 1, suffix 1, suffix 2, suffix 3 nth one and you can find out that the sum of the squares of these errors and that that is exactly what I have been talking about; find it for the nth plus 1, nth plus 2, nth plus 3 till whatever observations. So square them up, sum them up. So, that error should be minimum.

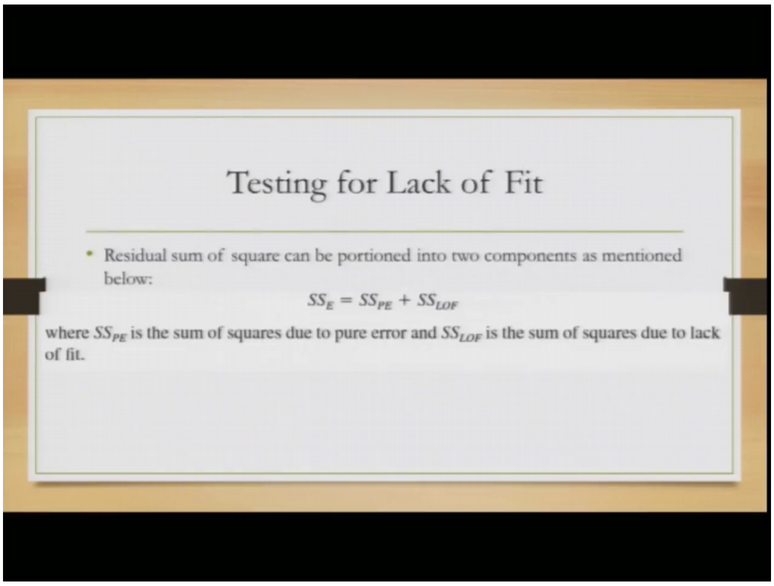
(Refer Slide Time: 05:50)



Thus this predicted sum of the squares use uses each possible set of n minus observations and we can find out using the observations in such a way then an estimate data set can be utilized and every absorption and turn is used to form prediction data set.

It would initially seen that calculating this value requires fitting n different regression models; however, it is possible from the reserve there is a single least square models can be utilized. So, you can utilize that and find it accordingly the sum of the squares and predict it accordingly.

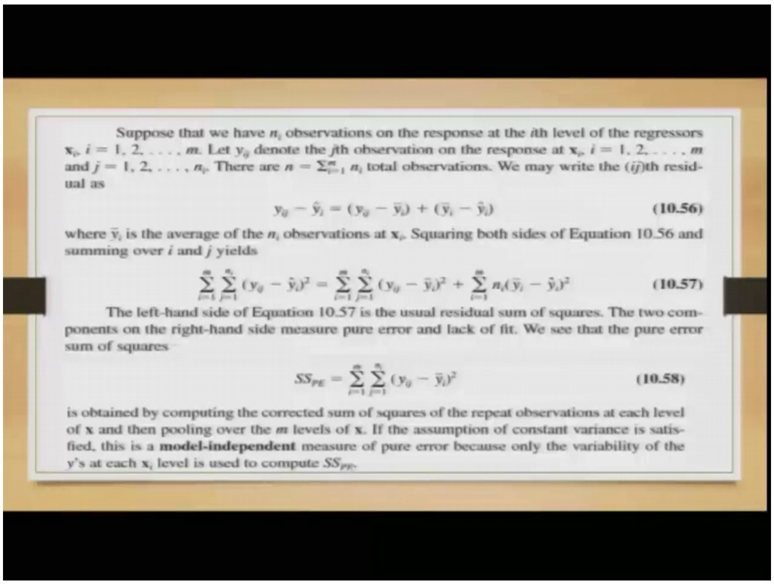
(Refer Slide Time: 06:25)



Residual sum of squares can be portioned into two categories; one is basically sum of the squares to the pure error and sum of the squares due to lack of the fit; so obviously, when you are trying to basically do that so in a regression coefficient model very simply; consider that you have a prediction of betas so once you find out they would be error there then once you use this betas to predict the y hat they would be error here they are basically there are two sets of errors.

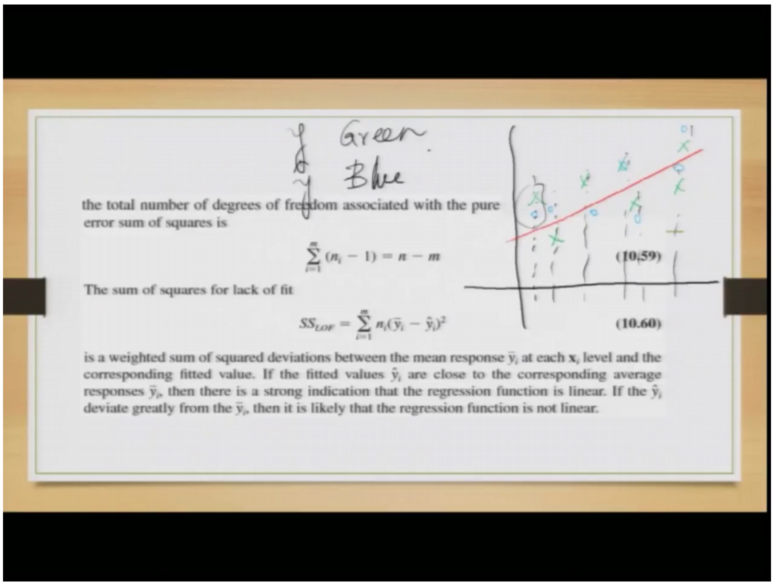
One is how good or bad your predictions are estimation I should use the word estimation of these betas are there and how good these estimated beta are able to predict the y hats. So, they are technically deviations or errors I would not use the word errors they are deviations or their differences in the actual one and the and the predicted value of the estimated value of two levels; one for the sets of betas, one for basically finding out the difference between the actual value of y and the predicted value of y. And even though it is not part of the course I would like to mention that some studies had been done by Zelnar using the balance loss functions and those basically will give you some idea on those lines.

(Refer Slide Time: 07:43)



Suppose we have n observations on the response of the ith variable so we can find out y i and denote the y i j. So, j th observation on the response for x 1, x i 1 2 3 4 till k and js are basically outside this n set of observations which you have. So, based on that you can find out the errors and once the errors are found out you can find out that the sum of the squares of the, of they were some lack of it and some are errors due to the fit. As I mentioned finding out betas would have an error finding on y would have error we can find it out accordingly.

(Refer Slide Time: 08:25)



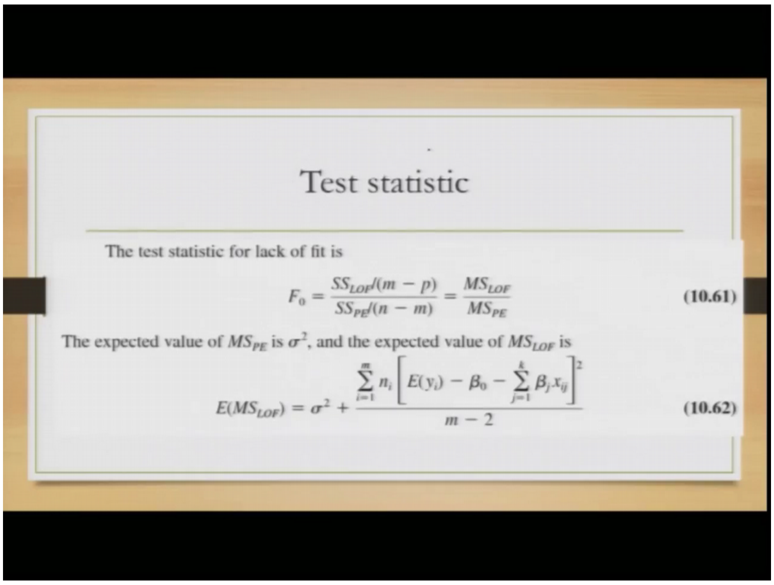
So, this sum of the squares of the square lacks of fit would be given by n i; which is the observations which you are going to take and the difference between the the mean value. So, what do you have actually so you have the best fit line; so from there you are trying to basically find out what are the errors.

So, y hat bar and this is the line which you have and this y hats are the predicted value. So, technically there are two sets. So, one is let me use this is easier for me to explain and for you to understand. So, this is x and y in 2 dimension; it can be in 3 dimension also; consider this so y is going along towards the roof, x 1 is coming towards me from one point and x 2 is going from my left to the right. So, it can be higher dimension also. So, base with now let me take a color; this is y hat these are actual y; actual y means y and these are the predicted y. So, this is the pen and a dotted one let me check. So, this is the first reading, this is the second reading, third reading, fourth reading, fifth reading, six th reading, nth reading.

For anyone you see the green and the blue. So, green is the actual y green and the circle one is basically hat blue. So, based on that you can find out one of the best fit it is. So, how close or far you have been able to predict. So, if you remember I mentioned find out the betas find out the errors there, find out the y s, find out the errors. So, you have basically trying to differentiate the errors.

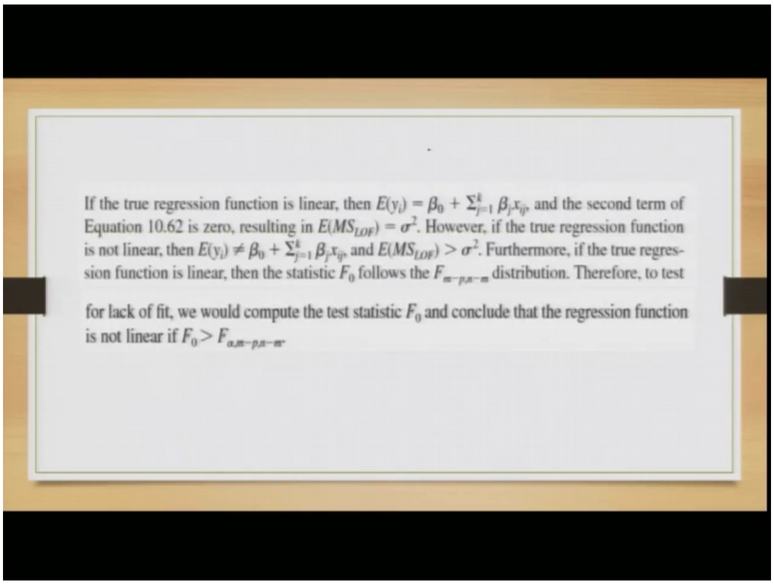
So, the sum on the last square of the lack of the fit is the weighted sum of the squares deviation between the mean response y bar at each level of x i and the corresponding fitted value the fitted value y hat are close to the corresponding average then the strong indication that the deviation function is linear. If I had deviate greatly from y bar then it is likely the regression function is not linear and you have to take decisions accordingly; whether the model is right or wrong.

(Refer Slide Time: 11:25)



The test statistics is again it come back to the sum of the squares of best fit and the error part find out those ratios; that means, sum of squares divided by the degrees of freedom and find out those ratios will give you their value under h naught you reject or accept h naught accordingly. And the errors under and this mean square for the best fit errors, predicting best fit traders, not predicting can be found out accordingly depending on how many such betas did you have.

(Refer Slide Time: 11:56)



If true regression is of a is linear then the expected value of y should be equal to beta naught plus beta 1, x 1 1, beta 2, x 2 1, beta 3, x 3 1. So, the first number if this is the variable number the next number is the reading number. So, when I am using if it is n plus 1, n plus 2, n plus 3 whatever it is; however, the true regression function is not linear then obviously, in that case the expected value would be either lower or higher for the more if the true regression function is linear then the F tests statistics and the h naught would find would follow a certain distribution and its degrees of freedom would be m minus p and n minus m.

So, m minus p; p is basically remember k or k plus 1 and n is the total reading which you have; m is the m as in mango is the new sets of observations which you are taking. Therefore, to test for a lack of fit we would compute the test statistics F naught and conclude the regression go function is the either linear or non-linear. So, before considering this experiment of random factors so we will basically try to wrap up the overall content on the course.

So, initially we did discuss of what is a very simply what is a random variable, then we considered that considering a random variable how what are the moments; that means, what are the dispersion concepts? What are the average tendency, average tendency can be arithmetic mean, geometric mean, then harmonic mean and they can be used to find out the than the height or the weight of a group of students trying to find out; the average speed of a car trying to find out the interest rate, flow of fluids during in channels.

Then we consider a different type of the dispersions can be apart from the variances you can have, different type of standard deviations with respect to median also. Median is basically a central tendency; where it divides the whole distribution into two equal halves of with probability 0.5, 0.5; then later on we consider different type of discrete distributions and continuous distributions in that in the discrete case we considered the uniform discrete case, Poisson distribution and so on and so forth.

In the continuous case we discussed the norm normal distribution exponential distribution and obviously, we laid more stress on the normal distribution because based on that we considered that three important sample distribution being chi square, f distribution and t distributions and chi square would basically be the relationship something to do with the second moment and t and m z would basically be something to with the mean value and f was basically something to do with the ratios of two different samples which we are going to take.

I mentioned that considering c t 1, c t 2 producing two different products of the same type and you want to basically find out the variability. So, you can use the f distribution. Then we went into the ANOVA table and ANOVA table I mentioned time and again I have been laying more stress in the last 10 or 15 lectures that your main emphasis is basically to find out the table where in all the variables would be there then you will have the sum of the squares and third column you have the degrees of freedom, in the fourth column you have the mean square values, the ratios of the of the total sum of squares by the degrees of freedom, then they have statistics and then depending alpha or p value will basically comment whether you support h naught or reject h naught.

We did discuss in different type of model of the fractional factorial models whether the variables were basically attributes or continuous variables and they would be of type 2 to the power k or of third of third designed, fourth design depending on how you want to basically partition that; we consider different type of fold models we consider different type of the assumptions where sigma square was considered to be fixed with respect to time; we did not consider it, but the concept of the matrix like this example came time and again in the regression part.

That we did not discuss the points where the variances were changing and the mean values could change, but you would only consider them in the linearity model where betas was 0 or not 0. Then we considered the fold the semi folds and then we consider the block designs and the how the block design is going to be done in such a way that you basically have the variables divided into blocks as the prediction level which was fine and they were on the on the on the highest values.

Values means that you want to basically reduce the errors part and basically try to predict as fast as possible; then we spend a lot of time about that if the factorials models can be 2 to the power k, 3 to the power k and this 3 or 4 to the power k would basically be the level of difference differentiation of any variables like it can be in the level of plus 1 0 minus 1.

It can mean the level of plus 2, plus 1, 0, minus 1, minus 2 so obviously, in the first case you will have you will have 3 to the power k in the next case you will have five to the power k it can be accordingly an even number also then we consider the different type of regression models the concepts of betas the assumptions which was there and the normal differ xs, normative ys, normative of the error terms, expected value of the errors being 0, variance being fixed, covariance existing between the xs and the errors was 0; obviously, they have to be. Then we could discuss about the rank on the matrix and basically considered the combination of the regression models with the ANOVA models.

This has been our technically are long intense course and I am sure that you have been able to tackle these assignments and obviously, he will definitely be able to do the exact final examinations in a much better way considering that we have are tried our level best to convey the concepts and I am sure considering the level of confidence which all of you have in within yourself you will do very well and for any queries whatever the we have been able to a best level of satisfaction for almost all of you.

So, as mentioned that for the regression models we considered the sum of the squares and we also consider the sum of the squares had a predicted value which was basically with r square. So, r square and adjusts the r square also consider the standard format where the degrees of freedom were important and degrees of freedom I did mention time and again what they work we have as I was mentioning also that, we have tried our level best to answer the queries through the forums and any for any help any advice you need.

For this TQM-2 lecture you are most welcome to get in touch with NPTEL office, main office in IIT Madras and we will we will from our side do our level best to answer the queries accordingly and again I am saying the books with they have been many queries about the books and the slides they are all taken from Montgomery’s books which is may be costly, but there are very cheap editions of Montgomery available in the market and or in the library also you can utilize that.

May be the concept which were considered maybe have been covered in the book in depth, but do not worry about that lay only stress on the parts you have studied or maybe you need not go into though all the detail pass you can utilize this book and this lecture for TQM-2 the best possible benefit for all of you. To do good in TQM-2 and also have an understanding of the design of the experiments which is very important for trying to understand; many of the decision making analysis from the point of view of statistics.

With this I will end this course and I have a very nice day and best of luck for all of you. I am sure all of you we will do very well in life.

Thank you very much.