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

Lecture – 39
Fitting Regression Models – IV

Welcome back my dear friends, a very good morning, good afternoon and good evening to all of you. This is the TQM-II lecture series under the NPTEL MOOC and I am Raghunandan Sengupta from IME department IIT Kanpur. So, we are in the penultimate stage of this TQM-2 and lecture series and we have two lectures left, which is the 39th and the 40th in the 39th I will try to cover more topics, in 40th I will try to basically wrap it up I will discuss what are the things we should lay more stress on in the 40th lecture. So, let us continue the discussion.

So, if you remember we were discussing that giving given a multiple linear regression; a linear one and which way if we check keep saying time and again. Our main hypothesis was basically to have them in a in the very simple linear format find out the beta hats, which and they should be unbiased, then try to find out the generally the plot of the errors with respect to all the x's; x's are the independent variable then have the q-q plots to find out the distribution of for the errors or it can be done for the x’s, y’s because intrinsically we are considering they are all normal.

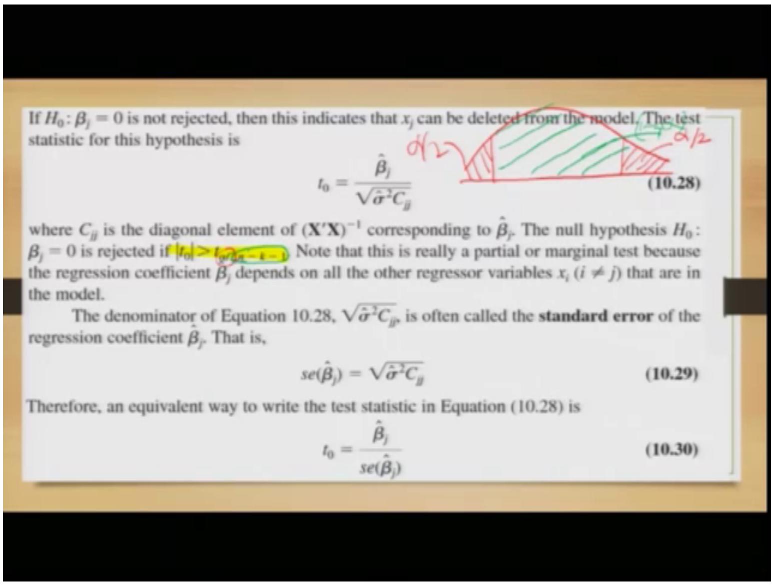
Then try to have a look at the covariance matrix or betas and then we went into how in if there are utilizing the concept of the regression with the fractional factorial models combining them; like 2 to the power k 3 to the power k. So, k is the number of variables or it can be attributes also. So, you can have a regression where some of them some of the x’s are discrete, sum of the x’s are continuous.

Then we considered that if they are orthogonals, how they can be done? Then on then the not orthogonal how the regression models can be done? And you can find out the beta hats. Then we also consider that if the folds can be done folding and trying to find out what is the overall relationship between x’s and the y's. Then we went into basically testing with the linearity holds at the fag end of the last lecture which is the 30th lecture we considered that if all the betas are 0; that means, there is no linear relationship between y and x. Because if beta 1 is 0 it means the rate of change of y with respect to x 1 is not there it does not exist.

So, that means, there is no dependence and then we will consider the further on discussion from now on. But you know all these things which I which I did mention is that at the end of the day your main focus is to find basically have a table of the anova which is analysis of variance. Where in the first column I am again repeating it will be all the variables and the attributes and the second column will be the sum of the squares then in the third column it will degrees of freedom, fourth column it will the mean square; that means, the sum of the squares divided by the degrees of freedom.

Then using the ratios of the or the mean squares or the ratios of the sum of the squares divided by the degrees of freedom will give you their statistics and you will basically have enough value of alpha, level of confidence or considering the p value you will comment whether your hypothesis null is validated or not. And then where we would basically consider as things are in the subsequent lecture which is which we are going to cover to it today.

(Refer Slide Time: 04:10)



Now, if H naught which is the null hypothesis beta j is 0 is not rejected, which means there is some linear relationship. Then this indicate that that x j can be deleted from the model. So, if it is not rejected we will and consider not consider depending on what ate actual hypothesis is. The test statistic for the hypothesis would be the t value. So, this t value or the t statistics would be beta hat j. Because beta hat j is basically the estimated value of beta and divided by sigma square hat c j j.

So, c j j are the diagonals of the X transpose X inverse matrix. So, that was one part when you are trying to find out the beta hats. So, beta hats would basically be found out as that in the long term, the expected value of the beta hats are exactly equal to the beta.

So, that is basically the characteristics of unbiasness. The t statistics are under H naught would be beta hat j and divided by square root of sigma square hat and c j j. So, they would be the corresponding columns of the principal diagonal of that. Because it is c j j it will be the values accordingly. The null hypothesis is H naught is basically if beta naught is j; it will be rejected, if the value of both because null hypothesis can be you are taking both sides it side it.

So, if it both sided obviously, it will be either on to the left of the light depending on which way we are looking at. The word what I am saying is that which way you are looking at is that whether the value of beta naught is equal to 0 or within the band range considering the interval estimation it is plus or minus; that means, plus some delta value with right of 0 and minus some delta value on the left of 0, if you are basically drawing the number line where 0 is the value where the null hypothesis basically is validated.

Now once you remember this concept of beta naught being equal to 0 can be changed as beta naught is equal to say for example, some fixed value where we consider the linearity relationship between the x's or x j's and the y's are constant and in another case a null hypothesis we can consider it is changing.

So that means, consider that a under H naught beta j is say for example, 3. So that means, the rate of change of y with respect to beta with x that corresponding x is 3 units which is the regression coefficient and on the null hypothesis it is 3 and on the alternative hypothesis is not equal to 3. Then also the interval estimation would be done accordingly where it will be 3 plus some delta 3 minus some delta you will basically consider that hypothesis is satisfied outside this interval is not satisfied.

Now this interval plus delta and minus delta which I mean mentioning would depend on the level of confidence and the distribution you are taking. So, it will be t distribution with corresponding values being there what are those values? I am going to come to that.

So the null hypothesis be beta naught is equal to beta j under the null hypothesis being rejected would be true if this is true; that means, it is greater.

So, now check here what I did mention the right hand left hand. So, I will just change the color. So, this alpha by 2 basically means the level of confidence. So, if you have I am just drawing the distribution very simply it is not normal it is basic consider is a t distribution. So these are the intervals. These values should be equal to alpha by 2, this value equal to alpha by 2.

And the region in between this color or this region would basically be equal to alpha. hence this oh 1 minus alpha sorry my mistake, 1 minus alpha. So, 1 minus alpha plus alpha by 2, plus alpha by 2 the total area is 1; as it should be. And finally, this n minus k minus one is basically the degrees of freedom depending on the experiment we are doing.

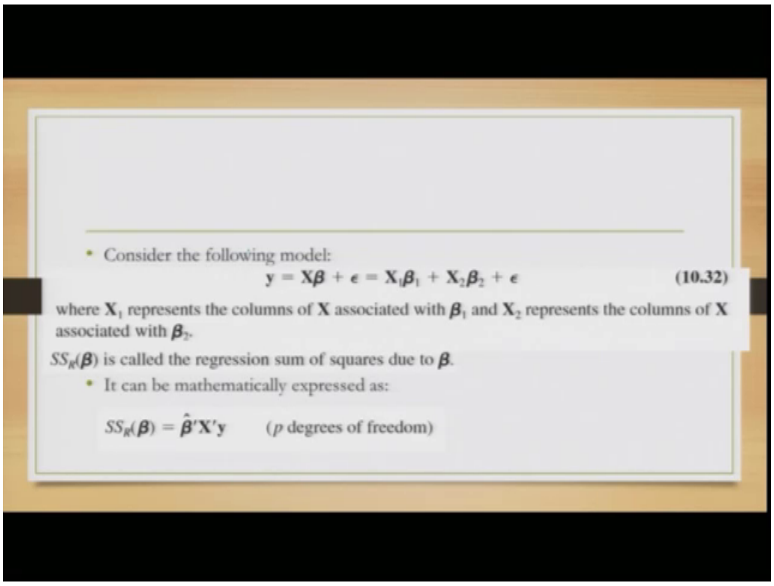
So, n is the total number of observations, k if you remember is the number of variables depending on which y can be estimated and minus one is basically you are losing the degrees of freedom for the for that was one. Note that this is really a partial marginal test because the regression coefficient beta j depends on all the other regression coefficient which is true. Because beta j would depend on beta naught beta 1, beta two 2, 3 till beta j minus 1 then beta j plus 1, beta j plus 2 so on and so forth del beta k.

On other regression coefficients ma I should basically mention not depend on beta naught till beta j minus 1 and beta j plus 1 till beta k, I am saying they should depend on the x values. So, x 1 to x j minus 1 from then x j plus 1 till say for example, x k. So, so let me continue reading it. So, the regression coefficient beta j depends on all the other regression variables x i; i is not equal to j that are in the model considered; the denominator of this equation which is the square root of sigma hat square c j j is called the standard error of the regression coefficient beta j because it is basically are trying to multiply the variance with the multiplying factor.

So, the standard error so s e is the standard error I will repeat I have mentioned it very clearly in the last class and also I faintly remember I did mention in the 37th class also. So, basically it means that standard error is the standard deviation of the sample. So, which is the counterpart of the standard deviation which you have from the population its counterpart in the sample is basically standard error.

So, let me continue reading it standard error for beta j is equal to square root of sigma square hat c j therefore, an equivalent way to write this test statistics would be as given t naught is equal to beta j hat obviously and and divided by the standard error of be beta j hat. So obviously, you are trying to normalize it and find out the standard the tests statistics.

(Refer Slide Time: 10:37)

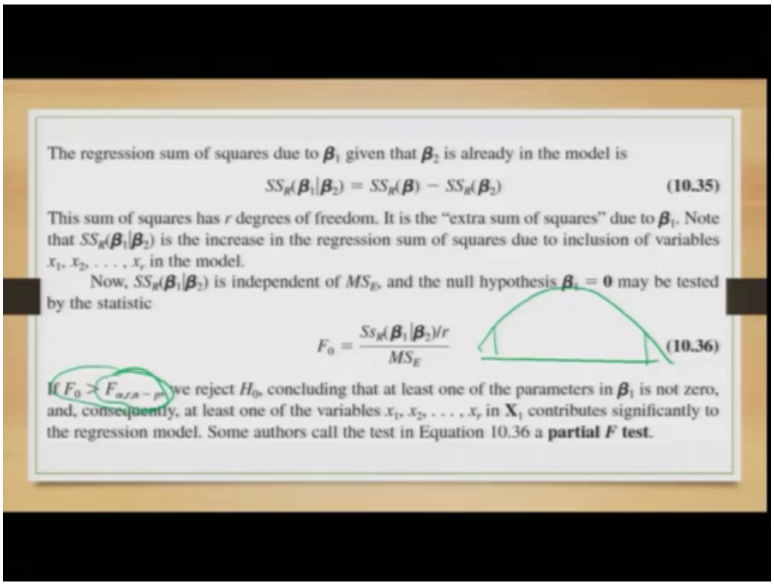


Now consider the model as given. So, you have basically y is equal to so there is no beta naught so is equal to beta 1 X 1 plus beta 2 ma X 2 and plus epsilon. Now the word which I mentioned there is no beta naught technically see the equation they are all bolds; bolds mean they all are matrices. So, let me if I read it will become clear to you where X i represents the columns of X associated with beta i.

So, basically have been able to divide the whole matrix into two parts need not be equal. So, X the corresponding value of X where and those that is X i, X 1 is a matrix is associated with the beta 1 vector and X 2 which is the matrix is associated with beta 2 vector. So, the stand this is the sum of squares are for the residuals which is beta is called the regression sum of squares due to beta and obviously you will have sum of squares coming considering the error part where the suffix will be capital E

It can be mathematically expressed as sum of the squares for the betas being equal to beta hat transpose. So, it was basically a vector it will now become column vector will become row vector multiplied by X transpose into y. So, all these are y is basically of size n cross 1. Similarly your X value would be of the size of k into n because then in that case X transpose y would basically now become k into cross 1 and so on and so forth.

(Refer Slide Time: 12:20)



So, if you basically do the multiplication accordingly for the matrices. The regression sum of the squares due to beta 1 and beta 2 k is already given in the model. So, given the sum of the squares for beta 1 given beta 2 or beta 2 given beta 1 would basically be given by this formula; where some of those squares of betas minus sum of the squares are from coming out from the beta 2 would be the sum of squares for beta 1 given beta 2. And if you replace 1 and 2 interchange them it will be exactly the same.

That means sum of the squares of beta 2 given beta 1 would be sum of the squares of beta minus sum of the squares of beta 1 the sum of the squares as r degrees of freedom; it is an extra sum of the squares due to the beta 1 which is happening. So, k the number of variables which are independent had been divided into r and k minus r.

So, whether r is for beta 1 or r is for beta 2 is basically a nomenclature and just a idea method how you are trying to do it, but remember that this matrix multiplication should be commensurate in the sense there are some and the rows and the columns should be such that the matrix multiplication concept stands true.

Note that the sum of the squares of beta 1 given beta 2 is the increase in the regression sum of the squares due to the inclusion the variables x 1, x 2, x 3 till x r. So that means, in a one case I am considering there to be added to the models and then what is the increase in the standard errors or sum of the squares in other case I am excluding them. So, what is the increase in the sum of the squares? Now sum of the squares are give beta 1 given beta 2 is independent of the mean squares and thus the null hypothesis where beta one is equal to 0 may be tested by the statistics given of f statistics, given some of the squares of beta 1 given beta 2 divided is by its degrees of freedom divided by the mean square for the errors.

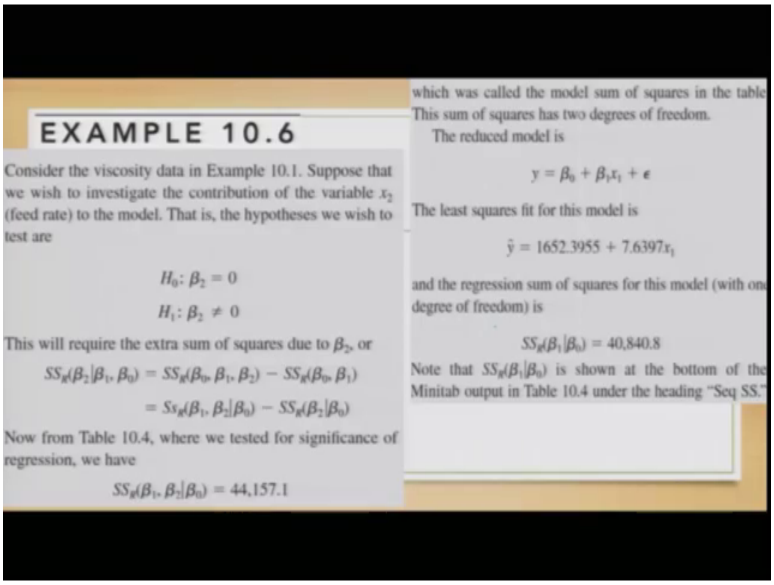
So, the errors also you have the sum of the squares divided by the corresponding degrees of freedom. So, you are taking the ratio of the sum of the squares divided by their degrees of freedom similarly in the denominator you also have the sum of the squares divided by the degrees of freedom you will give the f statistics under the null hypothesis and then you will test depending on the value of alpha and the p value whether you basically reject or accept H naught.

If F naught which is under null hypothesis is greater than that r f suffix alpha again I am repeating is the level of significance r is the degrees of freedom and minus p is basically the total dig. So, in F values you basically at you try to take the ratios of two samples. So, in this case one sample has degrees of freedom r and another case the degrees of freedom is n minus p. So, p if you remember depending on the problem it has can be either k plus 1 or k only depending on whether you have beta naught yes or no.

So, so that can be changed accordingly. So if F naught is greater than this value what is this value I will just circulate, so this one. So, the here F alpha again I am repeating is the this value.

So, this is one sided that is why it is not divided alpha by 2. So, the in the in the in the other equation we had basically left hands right hand side and left hand side. Now you are taking only one sided; because either greater than or less than So, it is greater than F alpha r is the degrees of freedom coming from the addition n minus p is the degrees of freedom which is coming from the total we reject H naught in concluding then at least one of the parameters in beta 1 is not 0 and consequently at least one the variables x one to x r in X capital X 1 which is basically the matrices contribute significantly to the regression model. Some authors call the test as a partial F test because you are to basically trying to find out the overall effect of r or a overall effect of not coming for our depending on that you basically pass a judgement whether you want to accept H naught or reject H naught

(Refer Slide Time: 16:17)



Consider the viscosity model again, suppose that we wish to investigate the contribution of the variables x 2 and that is the hypothesis we wish to test is accordingly under H naught beta 2 is 0. Under H a or H 1 beta 2 is not equal to 2 to 0. So, this will require the extra sum of the squares due to coming from beta 2. So, we can find out the sum of the squares of beta 2 given beta 1 beta naught.

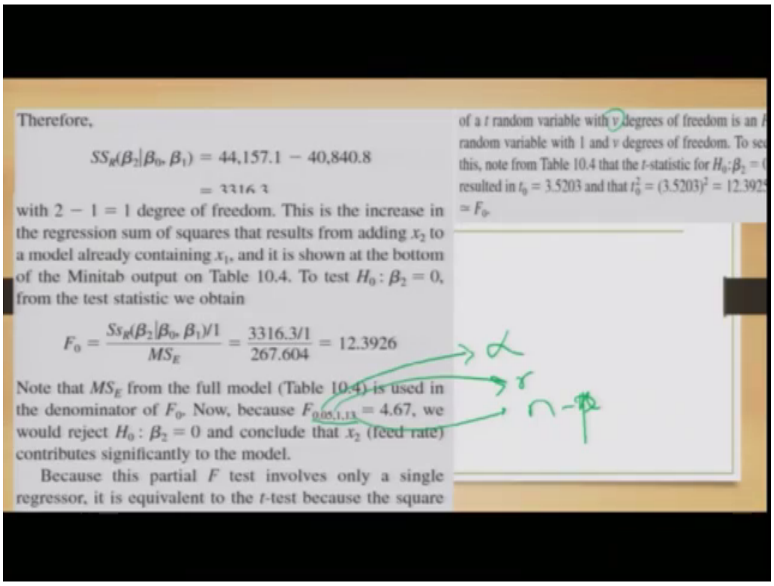
So now you have been able to differentiate and find out the matrices into two parts one coming from beta 2 one coming from beta 1 and beta naught. So now from this tables we can find out the sum of the squares of beta 1 beta 2 given beta naught is 44157 which is called this model sum of the squares in the table. This sum of the tables has two degrees of freedom depending on that the deduce model would be like this.

So, you have y hat now it is basically y hat because you are trying to estimate will be equal to beta naught hat which is 1, 6, 5, 2 that is 1652.39 plus beta 1 x 1 where beta 1 hat is equal to 7.6397 plus epsilon would not be there because this is hat hence we are trying to estimate so errors would come accordingly later on. And the regression sum of the squares of these models are given by sum of the squares of the of the of r given beta

1 for beta 1 given beta naught it comes out to be 40840. So, note that this value is shown in the bottom of the of the output and depending on this on the way you have basically been able to mention them.

So obviously, it can be sum of the squares of beta 2 given beta naught also depending on how you have been able to the so called partition the effects of r and n minus k.

(Refer Slide Time: 18:13)



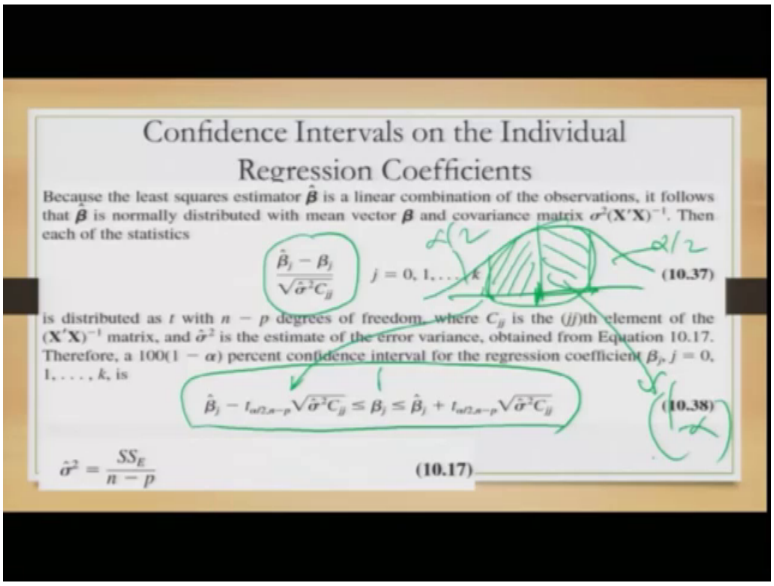
So therefore, S S R of beta 2 given beta 2 and beta naught is 3316.3. So, in this case you have degrees of freedom as 1 this is an increase in the regression sum of those squares there is results from adding x 2 to the model already occurred. So, you had x 1 you are now basically adding another variable x 2. So, if you keep adding what is the sum of the squares change.

Now and whether the linearity of the variables holds to test H H naught which is beta 2 is equal to 0 verses beta 2 not equal to 0, you can find out the F statistics that comes out to be 12.39. Known at the mean square of errors which is basically given in the denominators 267.604 is used, once you find out the values the F value F statistic based on which you will cross a judgment where the F naught is greater or smaller is 4.67.

Now, if you should note down these values. So, this is basically alpha this is basically the degrees of freedom which is r and this is basically n minus p; consider p p is basically k plus 1 or k depending on how you have been able to mention in a model. Because this is partial F's test involves only a single regressor, it is equivalent to the t-test because this square of the t random variable with chi degrees considered so it is written as v consider this is v degrees of freedom is an F random variable with 1 and v degrees of freedom accordingly.

So, here one is r and n minus k is v. To see note that that in the table we can find out H naught as beta naught is equal to 0 and you can test the t take t values comes out to be 3.55 and you can test using the t statistics and the values of t statistics are such that is almost equal to F naught. So, depending on whether you want to add 1 or take about 1 they would be equivalent to the t statistics, but if you can keep adding more than 1 then obviously in the F statistics value would change accordingly.

(Refer Slide Time: 20:31)



Now dif so this initially was about the hypothesis testing. Now if you want to basically consider the interval estimation of the individual regression coefficients because the least square estimates beta hat is a linear combination of the observation it follows that b hat is normally b hat beta hat is basically normally distributed with mean value of beta as mentioned according the concept of unbiasness.

And covariance matrix beam if you remember for the for covariance of betas so it will be a k plus 1 into k plus 1 or k cross k matrix with the principal diagonals are all the standard errors or the square of the standard square of the standard errors or the betas and the off the diagonal element are the so called covariance is existing between beta j and beta k or beta i into beta j where basically i and j change from 1 to k or 1 to k plus 1 depending on how many such variables you have.

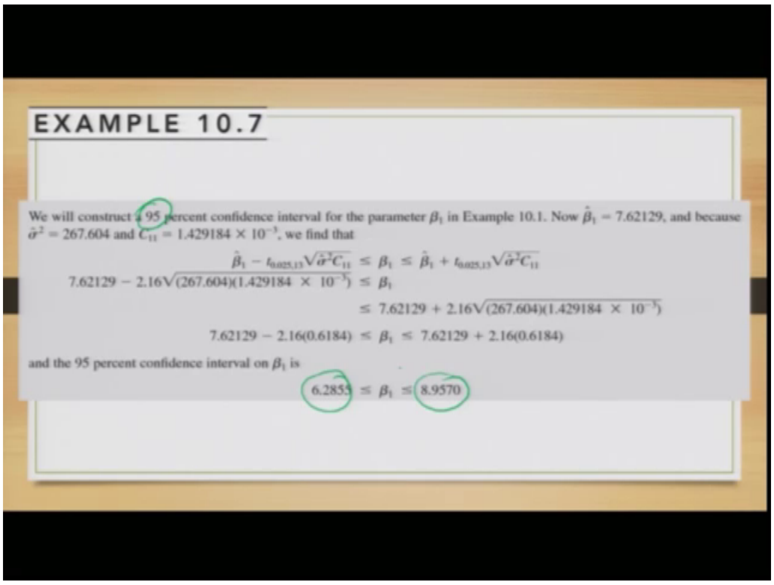
So, then each of the statistics would basically be that we want to find out the considering the normality holds. So it obviously it would become a very simple test of z value. So, in the z value what you have when you are con converting an x which is a random variable which is normal into z statistics or the standard normal distribution. So, it will be x in the new denominator you will you will have basically the square root of the variance or the standard deviation and in the numerator you will have basically that that random variable minus its expected value.

So, if you basically want to find a similar to that what I mentioned just look at this equation. So, it will be beta hat minus beta j. So, beta hat is the random variable beta j is basically is the expected value divided by square root of sigma square c j j which is if use mean remember the standard at standardized standard error.

So, it is distributed as t distribution with n minus p degrees of freedom, where c j j is the j j j th element of the on the matrix X transpose X inverse of that and sigma square hat is the estimate of the error variance obtained earlier. So obviously, this 100 minus 1 minus alpha is basically the levels of significance and based on that if you basically find out the interval; the interval would come like this.

So, this is basically the central value and plus minus a further interval in which you want to basically find out. So, the left value would be beta hat minus that that t statistics and on the right hand side it will be plus. Now note down one thing it is we are trying to find our interval. So, if it is interval obviously that value which you have so it will be alpha by 2 here alpha by 2 here. So, the whole area inside is basically 1 minus alpha. And n minus p is basically a degrees of freedom n minus k plus 1 or n in the bracket k plus 1 or n minus k would depend on number of variables which you have.

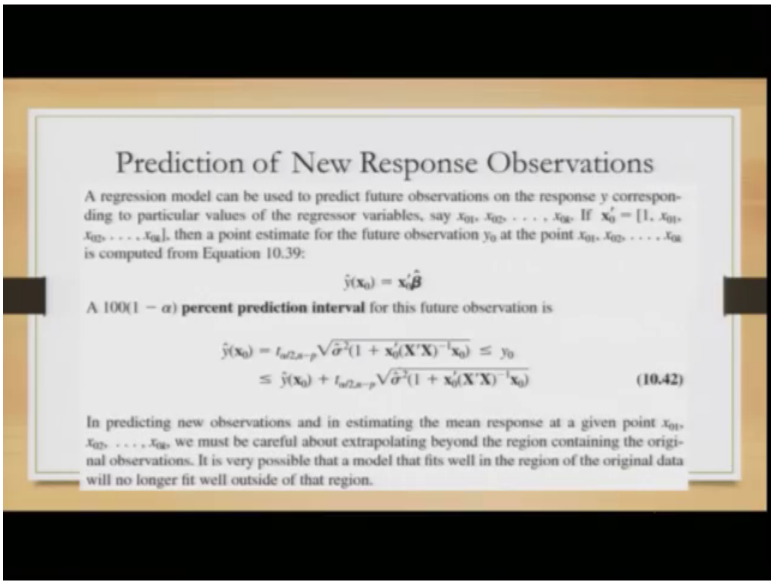
(Refer Slide Time: 23:47)



So, we can construct a 95 percent, 90 percent, 99 percent confidence in interval. Now beta 1 hat is equal to 7.62 and because sigma square hat is given as 267 and c 1 is given as 1.42 into 10 to the power minus 3. We can find out the confidence interval where the confidence interval for beta 1 comes out to be in between at a 95 confidence level it will come out to be 6.2855 is the left limit and right limit is basically 8.9570.

So, change 95 and the values of corresponding of the left limit and the right limit will change depending on that you can find out the very new limits for the confidence interval for betas depending on the level of significance.

(Refer Slide Time: 24:31)



Prediction on the new response observations; now consider a regression model it can also be used to predict future observation so obviously we have been mentioning time and again you want to basically find out y hat for the n th plus 1 y hat for the n th plus 2 so on and so forth find out the error difference for n plus 1 or n plus 2 or n plus 3 and if you want to double checks find out the sum of the errors they should actually be equal to 0 ok.

So, without repeating let me come to the problem a regression model can be used to predict future observations on the response y corresponding to particular values of the regression variables say x naught 1, x naught 2 so on and so forth in x naught k, so that means you are taking the first second third fourth the k th variable and their responses say for example, for the n th plus 1. Now in this case you will basically write X naught transpose as very simply mentioned is now in basically the scalar vector or a sorry is basically a row vector or a or a column vector where the first element remains as 1.

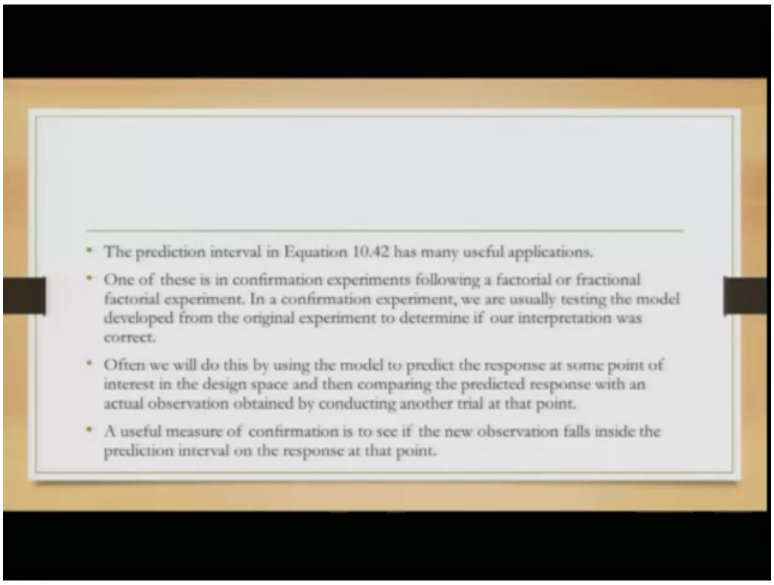
Because we are considering beta naught is there. So, hence point estimate or the future prediction model would be found out when we multiply if you remember correctly we will multiply what beta naught hat plus beta 1 hat in to x n plus 1 which is given as x 0 1 here plus beta 2 hat into x 2s x 2 which is the random variable its n th plus 1 value which is given as x naught 2 plus dot dot till the last value which will basically be beta k hat multiplied by x k th values and it is n plus 1 reading; which here is given as x naught k

So, if you I want to basically predict person 100 minus 1 minus alpha which is the level of confidence which you have person prediction interval for this feature observations you can find out that it will basically be y naught; y naught is basically the without the hat is basically y suffix n plus 1 reading. So it is left limit and right limit would basically be comprised of the values which you have for the using the t statistics and you can utilize that to find out that how the interval and depending on the level of confidence you have how the interval will change.

So, if we change alpha obviously, the intervals will increase or decrease depending on how you have considered that. In predicting new observations and in estimating the mean response to a given point set of values which is x naught 1, x naught 2, x naught 3 till x naught k we must be careful about extrapolating beyond the religion containing the original observation because if you remember and they would definitely be increase or decrease of the variance and increase and decrease on the trends so on and so forth.

So, how you are going to consider that that would be important we always be careful about extrapolating beyond the region containing the original observations it is very possible that our model that fits well in the region on the original data will no longer fit felt outside the region because as I mentioned for those points depending on change in the trend, change in the seasonality or change in say for example, pressure for of this example change in the viscosity, change in the pressure, change in the flow rate, change in the temperature and humidity and so on and so forth.

(Refer Slide Time: 27:56)



So, the prediction interval has would have many useful applications one of those is the confirmation experiment following a factorial or fractional factorial experiment. In a confirmation experiment we are easily testing the model develop for original experiment to determine if our interpretations were correct. So, given an in sample and out sample we will first test our model for the in sample and then try to utilize that model and that model for the out sample and find out the robustness of model.

Often we will do this by using the model to predict the response at some point of time of interest and the design space and then comparing the predicted response with an actual observations or it can be a useful measure for the confirmation. So, as to see if the new absorption falls within that in confidence interval such that we are quite confident that the model or the predicted values or the x's, y's and whatever model we have proposed is right. So, with this I will end the 39th lecture and continue some discussion the 40th one and then wrap up this course in the end of the 40th lecture.

Thank you very much for your attention. Have a nice day.