Data Analysis and Decision Making – I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 23

Welcome back my dear friends. A very good morning, good afternoon, good evening to all of you and this is the decision science Data Analysis and Decision Sciences. This is lecture number 23 as you can see on the slide and this total course on the NPTEL MOOC is for 12 weeks which is 30 hours; that means 60 lectures and each week we have 5 lectures each lecture being for half an hour.

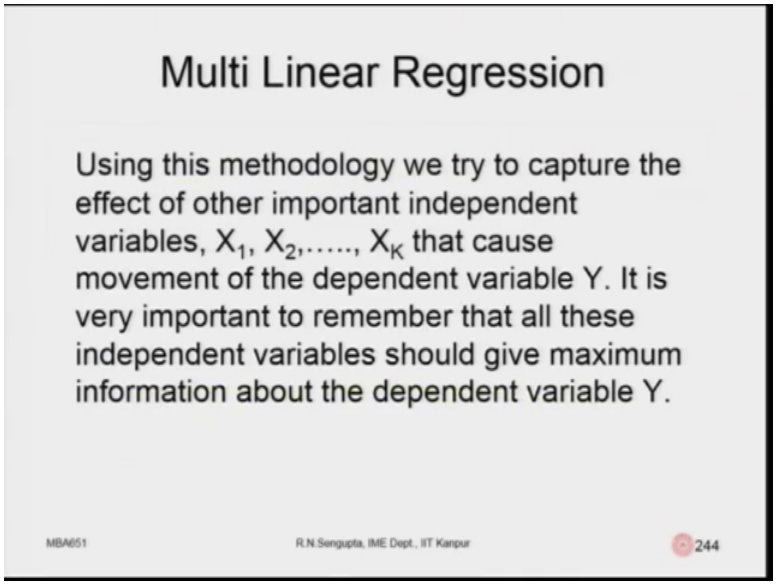
So, as we were discussing and by the way my I am Raghunandan Sengupta from the IME department, IIT, Kanpur. So, if you remember we are basically discussing the multiple linear regression concepts and or the simple linear regression concepts and I did discuss in details in some try to give you a feel that how does basically the assumptions hold. And, based on the assumptions I did also mention that given alpha and beta which are the parameters which you want to estimate what we will do is that we will basically find out the errors that is difference between the predicted and actual value square them up add them and differentiate that that error with respect to alpha put it to 0 with respect to beta put it 0 and basically solve them to find out the estimated value which is the hats.

And, then in the next period which is the nth plus 1 period we will try to utilize those hat values predict for the Y hat n plus 1 compare that with the actual value of Y n plus 1 and continue doing that. So, say for example, here the errors for the new time periods n plus 1, n plus 2, n plus 3 till the last term whatever the turn number of turns is and technically if the errors have been calculated using the actuation method which in give you the exact answers then the sum of the errors should be 0 because the expected value of the errors are 0 and obviously, the variance would be a fixed value which I mentioned as in the simplest case as one and in the next instance if we consider that it will be sigma square epsilon for suffix epsilon.

(Refer Slide Time: 02:24)

Now, for our problem which we just did what we if you plot the Y value which is the blue yellow line actual value Y hat is basically the red line and the errors if you remember I did mention the errors would basically have a mean value of 0 and a variance of sigma square. So, if the errors if you seen which is the dark blue one you will find out the expected value is coming out to be almost along the X axis hence the errors are basically 0.

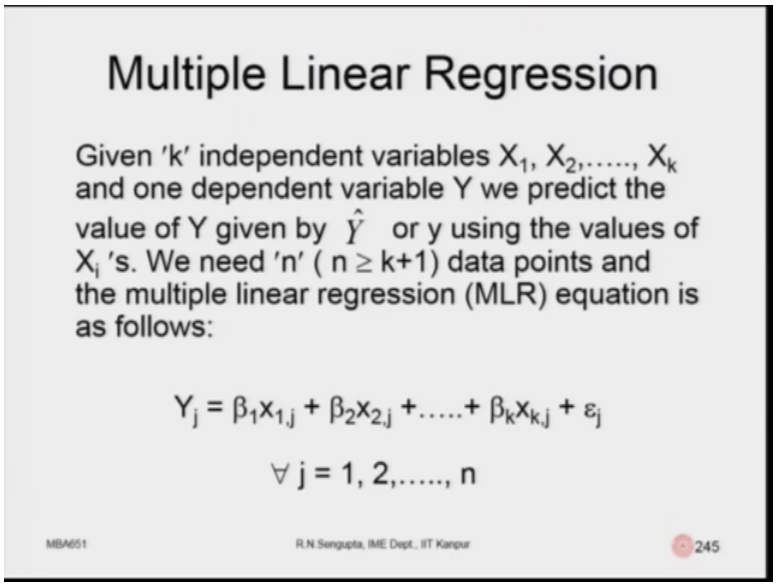
(Refer Slide Time: 02:46)



Now, using the same methodology same terminology same concept we will basically go into the realm of multiple linear regression. So, using this methodology we will try to capture the effect of other important independent variables X 1 to X k or X 1 to X p such that the combined effect which you have from by combining X 1 to X k would give you the best prediction value of Y which is the independent variable. So, let me continue reading it.

Using this method we try to capture the effect of other important independent variables X 1, X 2, X 3 till X k that caused movement of the dependent variable Y. It is very important to remember that all these independent variables should have the maximum information about the dependent structure of the variable Y. So, X 1, X 2, X k would give you the maximum information based on which you will basically be able to predict Y.

(Refer Slide Time: 03:40)



Given k independent variables X 1 to X k and one dependent variables Y we predict the value of Y is given by Y hat. So, basically using the we have Y you have different X's, you have the reading number 1 to n, you have the reading number 1 to n for all the X's based on that you find out alpha which is the fixed value if it is there. You find out beta 1, beta 2 till beta k provided there are X k values and then use this alpha beta 1, b 2 beta k all these are with hat values then once you find out they are the predict estimated values, using that you predict for the Y hat for the nth plus 1 value and then compare the

error which is the difference between the actual value of Y for the nth plus 1 reading and the predicted value of for Y at the nth plus 1.

So, we basically we predict the value of Y given by Y hat or small Y using the values of X's we need to have n such that n is greater than k plus 1. So, I will I will discuss something about why it is greater than k plus 1 and data points which are there and the multiple linear regression problem would be stated. So, let me first state the problem and then come to the discussion. So, you will basically have Y j or Y i whatever it is j is equal to 1 to n and the alpha value we are considering is not there and the equation is basically beta 1 into X 1 j plus beta 2 into X 2 j plus dot till beta k X k j plus epsilon j where the first suffix below X which is 1 2 3 4 so on and so forth are the variable number and the j is basically the reading number which is total reading number is n.

Now, why n should be greater than k plus 1? Now, if you see this equation you will find out that you will be able to solve this equation using simple matrix notations. So, let me write it down in the matrix notation format. So, basically what we will have is one Y I am writing it little bit spaced and why I will I will plug there plug in some values will understand it will be basically X into beta plus epsilon.

Now, if you see the number of readings the number of readings for Y is n. So, Y would be a vector column vector of size n cross 1. So, I am going to use the size write the size using different colors. So, the size is n plus 1. Now, if you see if you go to the epsilon value epsilon also has n plus 1 reading, so obviously, we will write the size of that also now technically Y is size n plus in n cross 1 epsilon is n cross 1. So, hence the multiplication of X into beta should also be n cross 1.

So, if there are k number of X's so, beta would be size k cross 1 and X would be size of n cross k such that when you multiply n cross k into k cross 1 you will basically get n cross 1. So, what actually the values or how would they look is this Y would be I will draw Y here Y would look like vector Y 1 till Y n size n cross 1. So, this thing this value this basically matches it whatever we were saying. So, this is satisfied. Now, if I come to epsilon on the right hand side epsilon 1 epsilon n this basically matches. So, I will try to basically highlight it this is what we will do matches this one matches.

Now, let us go through X and beta. So, let me first because space is limited. So, I will basically try to erase this ok. So, I was there. So, this I should also erase if I want to

basically denote beta here. So, this is done. So, let me first basically highlight X. So, X I will draw here. So, hope there is some space. So, X is the matrix which I said is cross n cross k. So, there are n number of rows and k number of columns. So, the first value would be X I will write the suffixes of the X later on, I am writing only the corner value. So, it will you will understand.

Now, if there are n number of rows so, I will denote the first value as the value for the rows just. Just a nomenclature you could have changed it how you could have changed it I will come to that later. So, this is 1 and this is the nth one, the first and the columns are denoted by 1 till 1. This will basically be the 1 into k this will be n into k. So, if you concentrate on the first cell 1 1 this is the first reading of the first the last one along the first column, this is the nth reading on the first the last one on the first row is basically the first reading for the kth and the last value on the last row or the last column would basically be the nth value for the kth.

Now, we could have so the nomenclature which you may find a little bit confusing is that why am I mentioning the suffix number as the reading number we will change it just give me 1 minute I will definitely change it. So, we could basically denote it as. So, this remains equation remains I will just remove this. So, what I will denote is this beta this is the transpose. So, beta transpose would be of 1 cross k and let me check if it is 1 cross k no it would not matter. So, I have to basically bring it as I will bring it the X 1 first the nomenclature now let me check I will I will just. So, this is n plus 1 and if I take X in the first place which is n into ok let us basically denote actual X. So, the nomenclature of the difference is only this.

So, if I denote actually X as a as a matrix of size n cross k. So, in this case actually what I will take is k cross not k n cross k as k cross n there I will use as X transpose beta X transpose is a value of n cross k. So, I should be using the blue color for our be n cross k and this would be k cross 1. So, n cross k into k cross 1 becomes n cross 1. So, only the nomenclature is seen here. So, actually X is a matrix of size k cross n and when you are trying to find out it you are basically writing X transpose.

Now, if I consider the nomenclature here. So, the first set I am just going to write in blue color in order to differentiate the first value, the second value, the third value and the fourth value. So, the fourth first value would basically be X 1 1 which is the first reading

first set. So, there are now n number of columns. So, the last one would basically be k is 1. So, the last one on this which is the second value would be 1 comma n which is the first reading nth reading. So, these are matching.

Now, you need to find out this. So, this will be the k value first reading and this will be kth value nth reading if I use a highlighter you know this highlighter is, yes. So, this value is here which is done let me use the orange this value is done which is here, let me use the red one even though that will resist make it very difficult to read this is the value which is here done and where is the other one, let me use the light green one. So, this is the value which is done.

So, you have basically seen the nomenclature being such technically this is would be giving you much better appeal where the first element denotes who you are and the second place would basically note at what standing you are; that means, the reading number. So, you can utilize the transpose of that and solve the problems accordingly, ok. The now I am coming back to the question why this is important. Now, when you want to basically solve it, so, I will remove all these things apart from this equation, this is not required. Let me remove this this this this this everything. Now, what we do? You want to solve them in the actual case and the errors in the expected values are basically 0.

So, when you find that the expected value these are all vectors called as whatever it is 0. Now watch here, when I want to basically find out the actual values average values of betas you want to find out I need to basically find out the inverse of X transpose pre multiply when I am hovering my electronic print pre multiply X transpose which is inverse hence it becomes an identity. So, it will be beta and on the left hands also as we pre multiply X transpose with it is inverse we will also pre multiply Y with the inverse of X transpose.

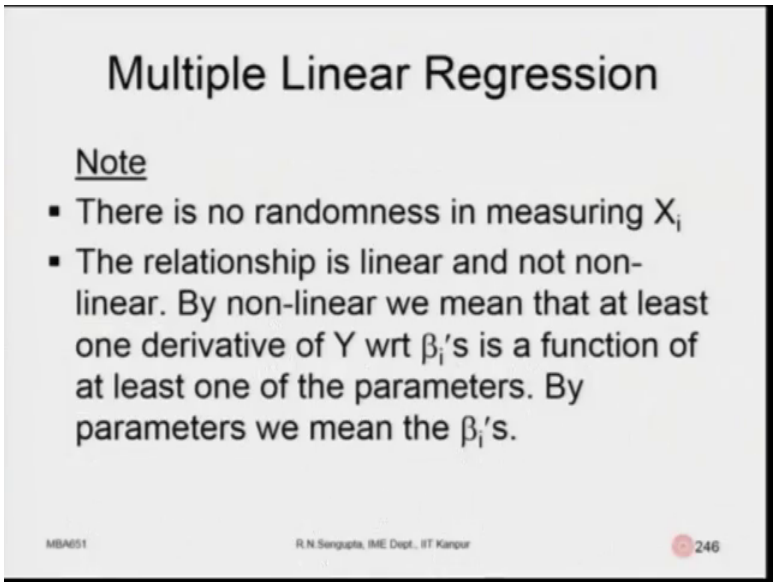
So, if you are basically trying to find out the inverse so, obviously, in that case the universe would exist point 1 and point number 2 if we remember in Gauss-Jordan equations when you solve the number of equations and number of variables should be such that the number of equations actually in the best possible case the number of equation should be exactly equal to the number of variables such that you have unique answer.

In case if you have a set of equation like this like 2x plus 3y is equal to 14 and consider your 3x plus 2y is equal to 15. So, you will solve it. So, it will basically eliminate X once find out X then put the X value and eliminate and find out Y. So, your job is done. So, in this case both the equations are independent on each other; that means, one cannot be expressed as a linear combination of the other. Now, consider the equation is like this.

So, if you see the second equation is a multiple of the first two times of that. So, actually solving and finding on unique values for X and Y is not possible which means the number of readings or the variables are such that the number of equations should be greater than equal to the number of variations or number reading number should be more than the variable number such that you will unique values for that. And, that is why when you are trying to utilize the concept as of the matrices this should be true such that the matrices would have a inverse and you are able to basically transform them and find out the inverse in order to basically solve and find out the unique values of the betas because betas are the estimate based on which you are trying to do the forecasting or the prediction or the multiple linear regression for the next time period which is n plus 1 n plus 2 and so on and so forth.

So, assumptions; so, if you remember in the first class I did mention the assumptions, I will again repeated maybe and that point of trying it was a little bit out of context now, but if you are doing this in the first class means when we started the multiple linear regression and this being and we had just started multiple integration I just gave a hint of the multiple linear regression and their assumption. So, I will again repeat it and this I am sure would make much better sense after conducting almost two lectures of half an hour each about simple linear regression and multiple linear regression.

(Refer Slide Time: 19:56)



So, assumptions are there is no randomness in measuring excise the excise are basically the independent variables the relationship is linear and non linear. So, what we mean by linear non-linear relationship I am going to come to that by non-linear equation we mean that at least one derivative of Y with respect to beta is a function of at least one of the parameters of the parameters by parameters we mean by betas.

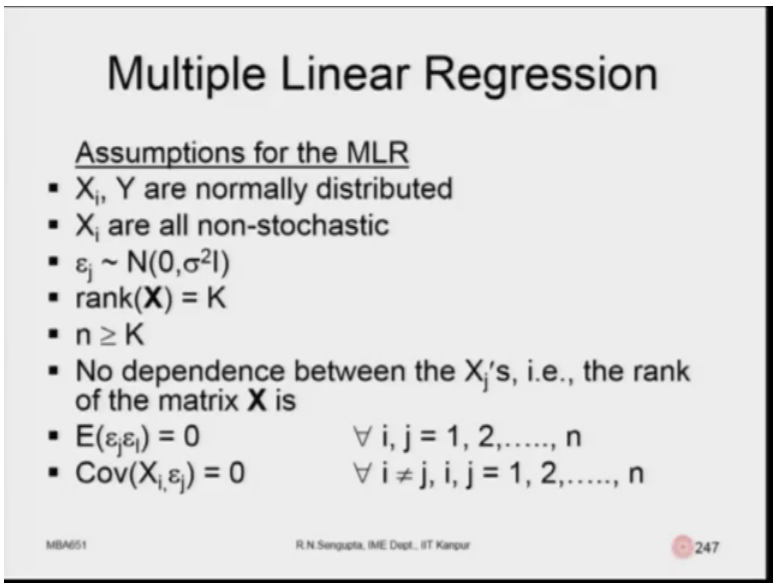
Now, consider why this statement is made. Say for example, if you had when you partially differentiate the equation. So, technically you put it to 0 and when you put it to 0 you find not be beta or b sorry for using the word b it is beta. Now, in this case if you had say for example, beta square, so, trying to basically differentiate the first time would not give you the answer because putting it to 0 what obviously, would mean the beta term is also already there. So, the partial differentiation for the terms of beta 1, beta 2, beta 3 which you are doing for the first time should be put to 0 in order to help you to find out what are the estimated values of the regression coefficient. So, they would not be any higher powers of betas it may be X square, it may be X cube, but. So, long it is linear if we mean main general like that what is actually stated by non-linear we mean that at least one derivative of Y with respect to the beta is still a function of the parameters itself.

Now, another thing is that when we are differentiating in this equation; that means, we are basically trying to find out the error, so obviously, errors would be the actual value of

Y minus beta 1 X 1 plus dot dot dot plus beta k X k. So, consider they are and square of that. They are k number of variables. Now, when you partially differentiate with beta 1 you consider beta 2 to beta k are kept fixed that is why you are partially differentiating putting it to 0 and finding out beta one hat provided all the others are fixed.

Similarly, when you partially differentiate beta 2 you consider beta 1, beta 3, beta 4 till beta k as constant and find out beta 2. Similarly, continue doing this. So, the point is when I mention as beta 1, beta 2, beta 3 are the partial regression coefficients actually it has that meaning that the others are kept constant other rate of changes are kept constant which means that if beta 1 is found beta 2 is found out it gives you the rate of change of Y with respect to X 1 or X 2 or X 3 provided others are fixed; that means, if I am trying to find out the rate of change of Y with respect to X 1 it will be given by beta 1, provided all the other random variables are kept fixed. So, here is the slide which we had discussed almost two classes back.

(Refer Slide Time: 22:58)



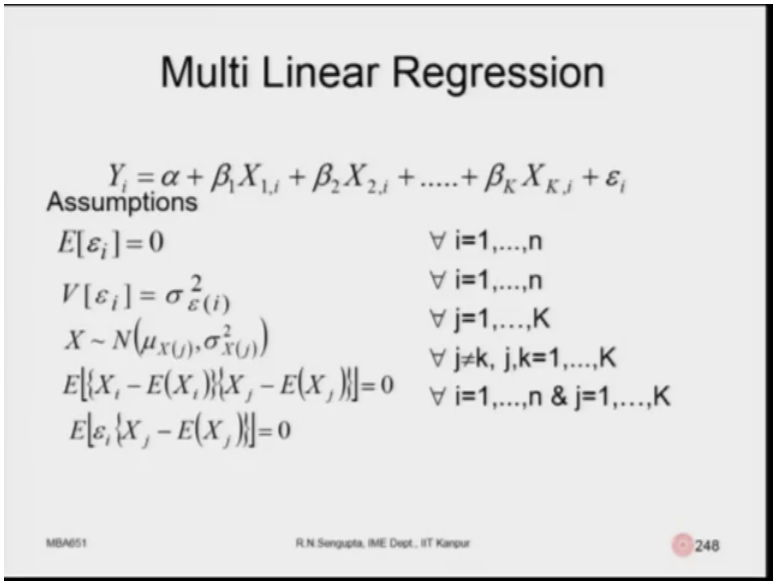
So, what are the assumptions for the multiple linear regression? The assumptions are almost exactly equal to what we have already discussed in simple linear regression. So, X 1 and X i's and Y's are all normal distributed point 1. So, obviously, if X's are normally distributed and in the simple linear regression I also said the errors are normally distributed and Y's are also normal distributed X i's are all non-stochastic the error terms has if you remember I did mention in a mean value of 0 and a standard

deviation as given. So, this is sigma square i; i is the identity matrix which you have the rank of X which is the X matrix is equal to k rank means basically you are able to using the rank you are able to basically consider all of them are independent. So, the rank is k; that means, there are k number of independent variables or else many of the rows or many of the columns when you write it in matrix form would be such that the rows and the columns can be expressed as a linear combination of the others.

The k value should be the n value should be greater than k and we are considering k as if you remember in one of the equations it was written k plus 1. So, in that case we have considering that alpha was not there in this case alpha would be there and if there is beta naught not beta 1, beta naught in means that there is no X's then beta naught would be subsumed as a matrix with a certain either as a row or a column depending on how your nomenclature for X's are.

No dependence between X's that is the rank of the matrix is as we have discussed. We will also consider and assume the covariances existing between the errors is 0. Covariance existing between the errors and the independent variables is also 0. So, that means, the errors between the error relationship between itself and the first term itself in the second term everything would basically be 0 that is not affecting each other.

(Refer Slide Time: 25:14)



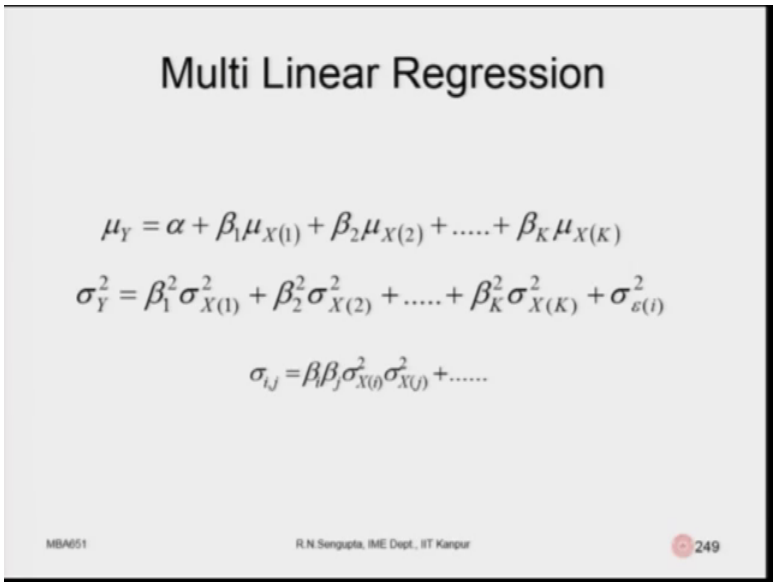
Now, if I find out if I write the equation, so, in the first case you basically have considering alpha is there or in many of the books it written as beta naught. So, Y i is

equal to alpha plus beta 1 into X 1 i; i is basic reading number, beta 2 into X 2 Y plus dot dot till beta k X k i plus epsilon and assumptions are the errors have a mean value of 0 which is what we have already discussed and if you basically plot it the errors would mean value be as if you remember we have done it.

The variance of the errors is given as sigma square and identity matrix, but we are considering it is constant it is not dependent on the reading number. The X's are each distributed with a certain mean and a certain standard deviation of the variance, so, the certain mean is mu X mu suffix X j; j is equal to 1 2 3 4 till k and sigma square X suffix j is basically the variances corresponding to the independent variable 1 2 3 4 till k.

Now, it also means that as I mentioned independence between the in the matrix form when you write it down between the x's; it means the covariance is existing between X 1 and X 2 or X i and X j is always 0 and the covariance is between the errors and basically the X's values are also 0. So, these are the last two terms which I mentioned these are this one.

(Refer Slide Time: 26:40)



Now, let us try to find out the expected value of the multiple linear regression. So, if you find out the expected value so, obviously, on the left hand side we have expected value of Y which is mu suffix Y which is right now what are the terms? Terms are alpha. So, there it is a constant term. So, the expected value alpha done, the second term is beta 1 into X 1. So, beta 1 is a constant you can take it out expected value X 1 is mu suffix one

which is done I am only talking about suffix one second third term is basically beta 2 into X 2. So, hence you find out the expected value it is beta 2 into X suffix 2 mu 2 which is done till the second last term which is beta k into X suffix k. So, that would be beta k will be constant. So, hence you find out the expected value of X k which is mu suffix k done and obviously, the last term which is epsilon the mean value is 0. So, it will be 0.

Now, come to the variance. Now, if you concentrate on the variance considering the assumptions the variance of Y will denote by sigma square suffix Y which is as written. Now, let us consider on the variances and the covariance of each and every term. Technically the variances of the covariance would be very small in number because the variance of alpha is 0 because it is a constant. So, this gone which is not there now from the second term to the last term the variances would be as follows and mark it very carefully the variance of the second term will be beta 1 square into sigma 1 suffix 1 which is right.

Second third term is basically third term in that equation which will be the second term here because the first term has already gone. So, this third term would be in that equation which is the second term here would be beta 2 square into sigma square suffix 2, till the second last term would be beta k square into sigma square suffix k and the last term would be the error of the variance of the error which is sigma square epsilon i, ok.

Now, the question is that what happens to the covariances? We have assumed the covariances between the X's them self, they are independent. So, it is vanishes covariances between the error and X's are not there hence is also vanishes and the covariance is existing between the fixed term which is alpha and the random variables; obviously, that would be 0. So, the covariance is existing between alpha X 1, alpha X 2 till the last one alpha X k and also alpha epsilon all will be 0 and finally, if I find out sigma ij basically it will be beta 1 into beta 1, beta 2 or beta i into beta j sigma square X, X i's and sigma square X j where beta 1, beta 2 are the partial correlation coefficients and sigma square suffix i and sigma square suffix j's are the corresponding variances of the ith and the jth X values.

So, with this I will end this lecture and continue more discussion over the multiple linear regression and the forecasting methods which we or the things which we have the smoothening techniques later in the related classes.

Have a nice day and thank you very much for your attention.