Regression Analysis and Forecasting

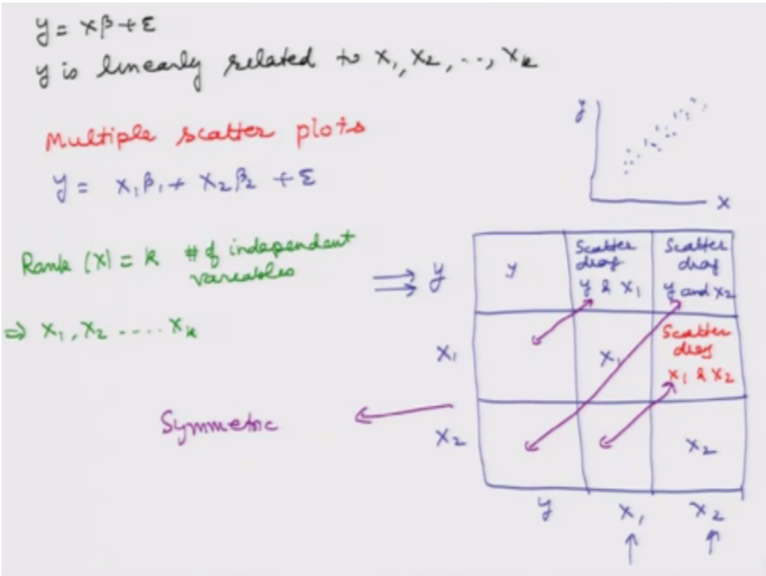
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology-Kanpur

Lecture -12

Estimation of Model Parameters in Multiple Linear Regression Model

Welcome to the lecture you may kindly recall that in the last lecture we had described the setup of multiple linear regression model like y=x beta+

(Refer Slide Time: 0:24)



We considered here that y is linearly related to x1, x2, x3 through the regression coefficient beta1 beta2 betak, before we try to estimate the model parameters a natural question comes that how do we decide that the relationship between y and x1 x2 xk is linear are not, so we take the help of a simple linear regression model in which if you recall I had a made a scatter diagram like this.

Where I have taken x and y independent and dependent variable and we had made a scatter diagram of the given observation and incase it is looking like this we can say that there is a linear trend and possibly the relationship between x and y is going to be linear, but now if you try see in this case we want to see the joint effect of x1 x2 x k on y is linearly related are not.

So in the earlier case of simple linear regression model we could work only with the 2dimensional figure, but in this case there is going to be high-dimensional picture and up to third degree possibly I can make a 3dimensional plot, but beyond third degree I cannot make a plot. In this case we tried to take the help of so called multiple scatter plot, multiple scatter plots are not difficult to obtain they are readily available in any software.

And in the case of multiple scatter diagram they are essentially trying to plot the scatter diagram among all the variables pair- wise. For example if I say if I have here a model something like y=x1 beta1+x2 beta2+ where I have got only two independent variable in this case the multiple scatter plot will look like this, this will be a sort of three by three diagram.

In which I had to plot y x1 x2 here and y x1 x2 here, so what really happen that in this section this will be a picture between y and y so there is no question of considering this graphic. Similarly here in this section the graphic is going to between x1 and x1 which has no meaning and similarly here the graphic is going to be between x2 and x2 which has no meaning.

But here this will give as a scatter diagram between y and x1, y is coming from here and x1 is coming from here, and here there will be a scatter diagram between y and x2, x2 will be coming from here and y will be coming from here. So, we try to look on this individual is scatter diagram and then we try to take a decision whether the joint relationship between y and x1 beta2 is going to be linear are not.

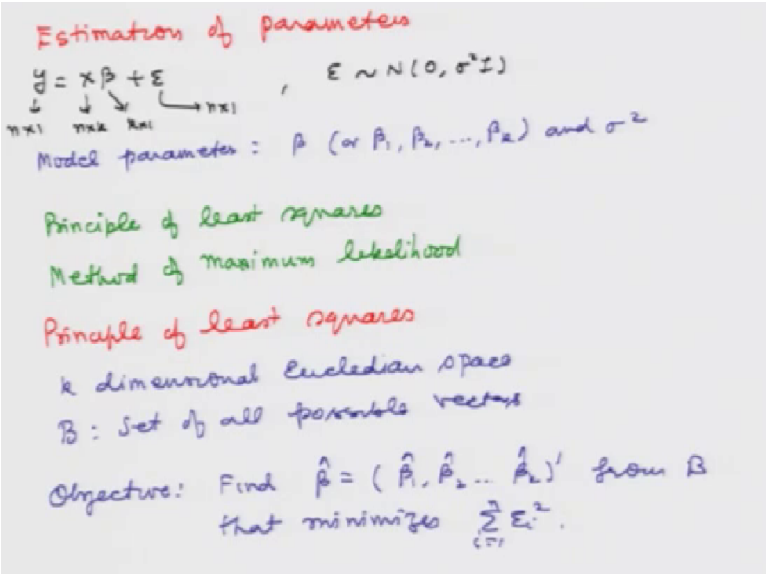
Well in a simple interpretation if I say if the relationship between y and x1 is linear y and x2 is linear then we expect the relationship between y and x1 and x2 also be linear, but there can be some more complicated situation so we will try to take up some numerical example later on and we will try to demonstrate that how do we take a decision in such situation.

So, one think we have to notice that this section will give us a scatter diagram between x1 and x2, what is this mean? I how would try to demonstrate it clear that how does it help us actually if you remember we had made an assumption that rank of x matrix is going to be k that is the same has the number of independent variable or number of explanatory variables, so this implies that all x1, x2, x k they are going to be independent.

They are going to be other linearly independent, so we need to ensure before we go for any linear regression analysis that whether these assumptions are verified or not so this multiple diagram will help us in diagnosing the issue that for example in this case x1 and x2 are independent are not, so incase if we see is a sort of linear trade or see any other type of relationship that would indicating the x1 and x2 are not independent.

In this multiple scatter diagram this is a symmetric diagram, this is symmetric, symmetric in the sense that whatever is the picture here the same picture will be here and the same picture will be here this is how using the concept and idea of multiple scatter diagram we decide whether the relationship between y and all x1 x2 x k is linear are not.

(Refer Slide Time: 06:53)



Now we come to another important aspect and we try to estimate the model parameter, so again just for the sake of understanding we are going consider here a model y= x beta+ epsilon, where y is a n cross1 vector of observations on study variable, x is a n cross k matrix of n observation on each of the k independent variable and beta is a k cross1 vector of regression coefficient beta1 beta2 betak and epsilon is a n cross one vector of random error components.

We assume that f silent is following a normal distribution with mean vector zero and covariance matrix sigma square i. So our model parameters in this case are beta vector or it is equivalency that beta1 beta2 betak and sigma square there are altogether+1 model parameters that need to be estimated to describe the model completely. So in order to estimate these model parameters we have two options.

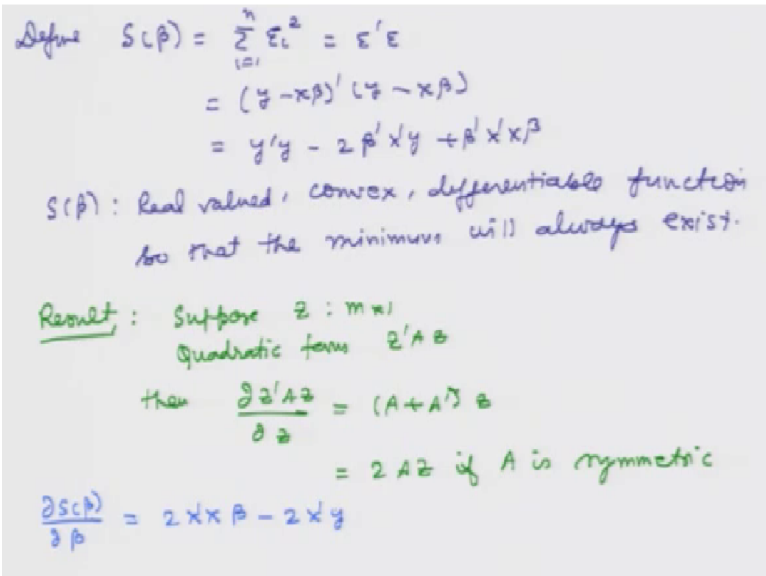
First think is this I can use the principle of least squares and second is method of maximum likelihood. Just like in the case of simple linear regression model we will try to use the principle of least square as well as maximum likelihood estimation in the case of multiple linear regression model also, so first we are going to talk about the principle of least squares. If you recall incase of simple linear regression model.

The model was y= beta0+beta1 x+epsilon and we had minimize the sum of squares due to random error summation i goes from 1 to n epsilon i square. Our objective was that we wanted to find out the value of beta0 and beta1 by minimizing the sum of square due to random errors, now in this case if you try to see this is not a two dimensional problem.

But there are parameters beta1 beta2 betak and sigma square, so obviously first we will try to concentrate on the estimation of beta1 beta2 betak and sigma square will be estimated later on separately, so if you try to see we are working here in a k dimensional Euclidean space. Let us try to define suppose capital b is the set of all possible vectors in this k dimensional Euclidean space.

So now our objective is that we want to find out the value of parameter vector beta such that the sum of squares of random error is minimize over is entire space, so now I can define my objective, my objective is to find say beta hat which is actually beta1 hat, beta2 hat say this betak hat vector from B that minimizes summation i goes on one to n epsilon i square.

(Refer Slide Time: 11:25)

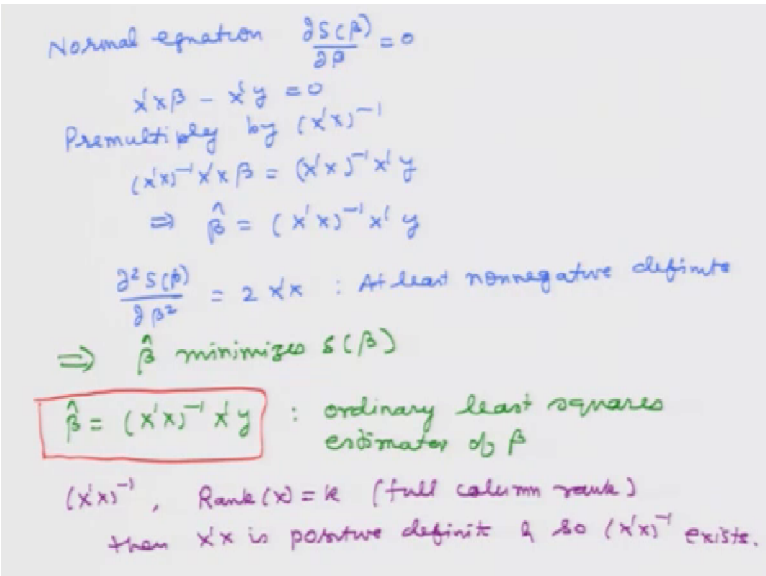


So now let us try to do it so we will define here a function say s beta which is something like summation i goes from 1 to n epsilon i square and this I can write down as here epsilon transpose epsilon where epsilon is n cross1 vector, now the same thing can be written as epsilon y- x beta transpose y minus xbeta and if I try to open this then we get here y transpose y -twice of beta transpose x transpose y + beta transpose x beta.

We assume that this s beta is the real valued convex differentiable function, so that the minimum will always exist. In order to find out the minima we are going to use here a simple vassal so before doing anything let we state this result, suppose z is a random variable of say order m cross one, so this is a m cross m one vector, now in case if I to try consider here a quadratic form something like z transpose a z.

Then partial derivative of z transpose z with respect to this z is A+ A transpose times here z and this becomes twice of A z if A is symmetric. So we are going to actually differentiate the function s beta and during this process we are going to use this result. So now if we try to differentiate s beta, partially differentiate s beta with respect to beta we get a twice of x transpose x beta - twice of x transpose y.

(Refer Slide Time: 14:12)



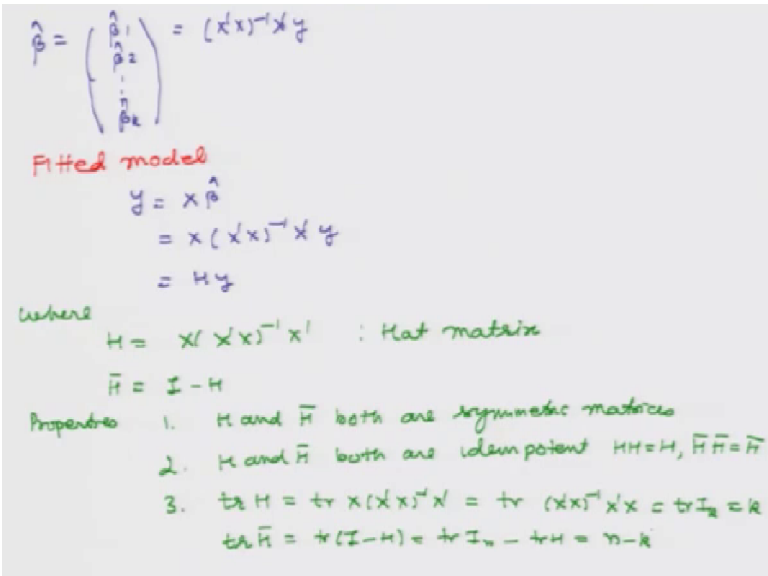
So from here I can obtain the normal equation by partial derivative of s beta with respect to beta=0. We have this equation here as a x transpose x beta- x transpose y=0, so what we try to do here that premultiply by x transpose x whole inverse, so we get here x transpose x whole inverse, x transpose x beta=x transpose x whole inverse x transpose y, so this finally gives us that beta hat=x transpose x whole inverse x transpose y.

Now we have to also ensure that this value of beta hat really minimize the sum of a squares due to random error which is s beta, so in order to do so we try to find out the second order partial derivative with the respective beta, and this comes out to be here twice of x transpose x. So this matrix is at least nonnegative definite, and so this ensures that beta hat minimizes s beta and then this now beta hat=x transpose x whole inverse x transpose y this is called has ordinary least squares estimator of beta.

One think we need to note here is the following that here we are going to use that term x transpose x whole inverse, how do we ensure that whether this inverse will always exists or not, so if you recall we had made an assumption that rank of x matrix =to here k, this was a full column rank matrix so then x transpose x is positive definite and so x transpose x whole inverse always exists.

So you can see here that whatever assumptions we had made they have some utility they are going to be use in the mathematical derivation of the results, and they will have some practical issues also that will try to address later on. Now what you have to always keep in mind this result that this ordinary least square estimator of beta and remember one think here beta hat is not a scalar, beta hat is a vector here something like this.

(Refer Slide Time: 17:39)



If you try to see here I can write this beta hat is actually beta1 hat, beta2 hat and betak hat, and so whatever you are going to get after solving this equation you will get a vector and the individual values of that vector they are going to give us the individual estimates of the elements of beta vector. So once I have obtain beta hat then I can obtain the fitted model the definition of the fitted model carries over from the definition.

That we describe during the simple linear regression model that we try to estimate the model parameters and we substitute the model parameters in place of the parameters and the model that we obtain that is called as a fitted model. In this case if you see we had a model y=x beta + epsilon and now we are going to replace beta y beta hat so this become our fitted model.

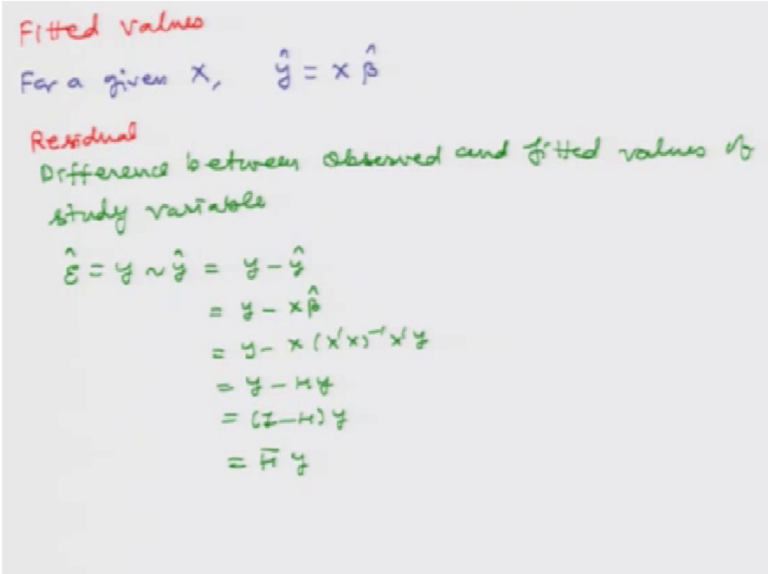
So this can further be written has something like this x transpose y, so here we will try to denoted by here hy and where h is x x transpose x whole inverse x transpose. This matrix has a particular name and this is called as hat matrix, and you will see later on that this hat matrix plays an important role in the linear regression analysis and h matrix as some nice properties also, but before that I can also define another quantity say h bar which is i-h, i is an identity matrix.

So both this h and h bar they have certain properties, first important property is that h and h bar both are symmetric matrices, second important property is that h and h bar both are inden potent so that means h into h=h as well as h bar into h bar =h bar. Third important properties that we are going to use are the trace of h, what is the trace? Trace is the sum of the diagonal elements of a matrix.

So if you try to find out the sum of a diagonals of h matrix then this is trace of x x transpose x whole inverse x transpose and we have a popular result in the matrix theory that trace of ab=trace of ba where a and b are the two matrices and their product is well defined, so I can write down this thing here as a trace of x transpose x whole inverse x transpose x and then this become trace of ik and ik is an identity matrix of order k by k so this trace is actually here k.

Similarly if you try to find out the trace of h bar this becomes a trace of say i-h, so trace of identity matrix of order n by n- trace of here h and this comes out to be n- k so these are some result that you will be using later on.

(Refer Slide Time: 21:53)



Okay, now after that let us try to describe how to we define the fitted values you may recall that in the case of a simple linear regression model we had obtain the fitted values by providing the values of independent variables to the fitted model. Similarly exactly on the same lines I can define the fitted values here as for a given x the fitted values are denoted by y hat and they are they are define as yx beta hat.

Next quantity we are going to define about the residuals, if you try to recall during simple linear regression model we had define the residuals as the difference between observed and fitted values of study variable exactly on the same lines I can definite here as a f silent hat which is y difference y hat and so more logically I can write down here say y minus y hat, so now we have obtain y.

y hat here is x beta hat and beta hat we had obtain as a here y minus x x transpose x whole inverse x transpose y, so I can write down here y minus hy and so I can write down here i minus h times here y, so this is equal to here h bar y, so this is a very convenient notation for us to use to present the epsilon hat which is our residuals. Now we stop here and in the next lecture we will try to investigate the properties of ordinary least square estimated of beta, till than good bye.