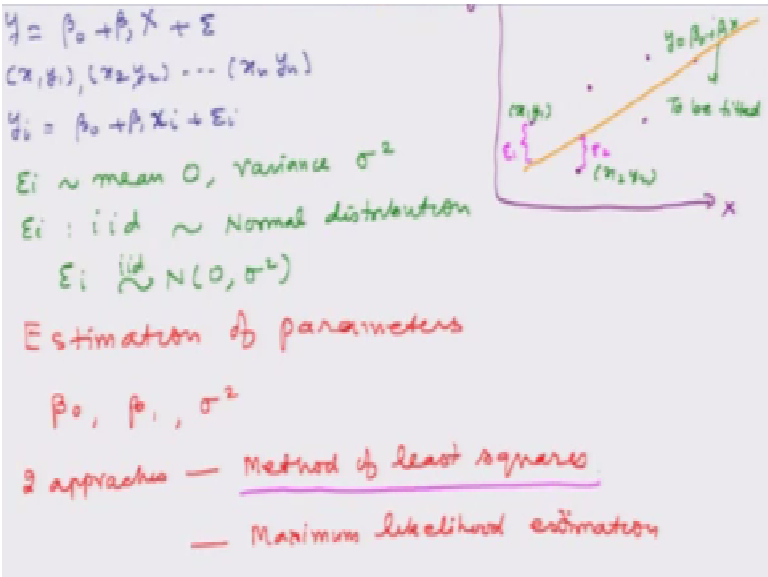
Regression Analysis and Forecasting
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology – Kanpur

Lecture – 04
Estimation of Parameters in Simple Linear Regression Model

Welcome to lecture number 4, in this lecture we will discuss how to estimate the parameters of a linear regression model, in the earlier lecture we had discussed that there are 3 parameters beta0, beta1 and sigma square, so if you try to recall in the earlier lecture we had taken the model y=beta0+beta1x+epsilon.

(Refer Slide Time: 00:32)



And we had obtain n observations say x1, y1, x2, y2, xn, yn and we assumed that all this observation they are going to satisfy beta0+beta1xi+epsiloni, this is the model that they are going to satisfy, and if you try to recall we had created this diagram this was x, this was y, and then we had observed the point something like this and so on and we wanted to fit here a line something like this.

We had given it a name say x1, this is my x1 y1 and this is my x2 y2 so this line is now in more technical terms this is the line which we want to be fitted and this is essentially the line y=beta0+beta1x. So in this case we also had assume that this epsilon i has got mean 0, variance sigma square, and we also assume that epsilonare IID that means they are identically and independently distributed.

At this movement I am going to make an assumption that epsiloni are IID and they are following a normal distribution. I can write down briefly that IID, epsilon are IID, following normal 0 sigma square distribution. This mean that all epsiloni they have been observed from the probability density function normal with mean 0 and variance sigma square, and we also assume that they are independent they all epsilon1, epsilon2, epsilon3, epsilon n they re mutually independent of each other.

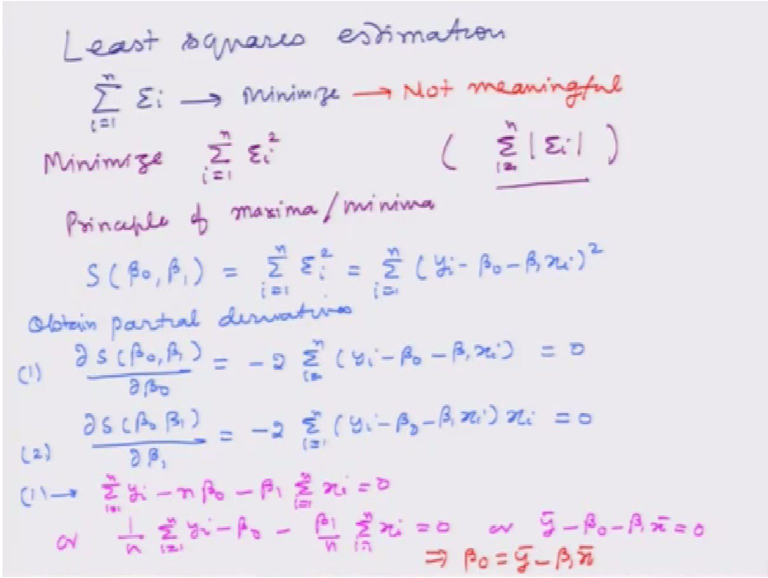
I would like to make here one note that when we are going for the least square distribution this assumption of normal distribution will not be of use there. When we are going for the test of hypotheses and confidence interval estimation then only this assumption of normality will be used, and later on when we are doing the maximum likelihood estimation in that case right from the first step we will require assumption of normality, so that you have to keep in mind.

Well, I will try to explain you as soon as I come to the maximum likelihood estimation and ordinary least square estimation. So under this setup now we try to estimate the parameters so our objective is estimation of parameter, and you have to keep in mind that there are three parameters beta0, beta1, and sigma square that we want to estimate.

Now I am going to use here two methods or 2 approaches, 1 is method of least squares and another is maximum likelihood estimation, first we try to understand what is this method of least square, now in this graphic if try to see we had said this is my random error involved with the first observation denoted has epsilon1 and similarly this is my epsilon2 and so on.

So if you try to see in very observation I have got some random error now principle of least square says that I would like to find out this line, this orange line in such a way such that this random errors are minimum and most of the point they are lying exactly on the line, so in the first case I try to use the method of least squares. So first let us try to understand what is the least squares estimation.

(Refer Slide Time: 05:28)



The principle of least square says that we try to find out the values of parameters in such way that the total error is as minimum as possible and most of the points are lying on the line. So if you try to see in this picture the random errors in the first observation is epsilon1, in the second observation this is epsilon2, and so on, so incase if you try to minimize the total error the total error is summation i goes around 1 to n epsilon i.

But can you really do it, if I say we have to minimize it, does it make any sense. We had assumed that some of the errors are in the positive direction that is above the line and some errors are in the negative direction they are line under the line. So if try to sum them up, some may be very close to0 and that will be indicating that my observation do not have random errors, that is wrong

So this idea does not work here, so this is not meaning full, so now how to do? Let as try to considered and let us try to minimize i goes from 1 to n summation epsiloni square. Now does this make any sense? Answer is yes. Why? Because we had face the problem earlier because some of the random errors were negative so once I try to square them the negative become positive and now I can easily minimize it.

Well, at this stage you can ask that once I am trying to convert my negative random errors into positive random errors then another option is that I can take the absolute value of epsilon , yes, answer is yes. You can also minimize i goes from 1 to n absolute value of summation epsiloni, yes, you also minimize the sum of absolute errors that is i goes from 1 to n, summation epsilon.

This is also available in the literature this is called as least absolute division estimation technique, but in this course we are not going to talk about it. So we will try to consider that we want to obtain the values of the parameters by minimizing some of squares of the random error. So the next question is how to minimize it? Well, I can use the principle of maxima in minima.

Let us try to use the principle of maxima/minima and try to obtain the values of beta0, beta1 and sigma square so let me write the summation epsiloni square has as a s of function of beta0 and beta1 i goes from 1 to n, epsilon square. This can also be written has a summation i goes from 1 to n, yi – beta0,- beta1, xi whole square.

The principal of maxima and minima says that we need to obtain the partial derivative of s with the respect to beta0 and beta1, put them =0 solve it, and then check using the second order derivative whether the solution gives us the maxima or minima, so exactly we are going to follow the same rule. So if I try to obtain the partial derivatives first.

So I try to obtain the partial derivative of this thing with the respective beta0 and this will come out to be - twice of i goes from 1 to n, yi - beta0, - beta1xi and next I tried to partially differentiate this s with respect to beta1 and this comes out to be summation i goes from 1 to n y i - beta 0 – beta1 xi times xi. And now I try to put it=0 put them =0 and I need to solve it.

So let we call this as equation number one and equation number two. If I try to solve this equation number one this can be obtained like as follows, once I open the bracket this gives me i goes from 1, 2 n summation y i - n times beta0 – beta1 times summation i goes from 1 to n xi put it=2. Or I can write down this thing here as a summation i goes from, 1 to n say y i – beta 0 - beta 1 over n summation.

i goes on here 1 to n xi= 0 or I can write down here ybar – beta0 – beta1 xbar =0. So solving this thing this gives me that beta0 = ybar - beta one xbar, but this beta0 can be known to us provided beta1 is known to us, but up to now we do not know the beta1 so now I try to solve this equation number two and let us see what we obtain over here.

(Refer Slide Time: 12:07)

Consider (2)

$$\sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\underline{\underline{\beta_1}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \boxed{\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = \hat{\beta}_1$$

Estimator of β_1

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \qquad s_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

estimator of β_0

$\hat{\beta}_0, \hat{\beta}_1$: Direct regression estimators

: Least squares estimators

Let as try to consider this equation number two and we try to solve it. The equation number two is summation i goes from 1 to n xi, y i – beta0 - beta 1 xi = 0, and now if you just try to open the bracket and if you try to solve it we get here beta1= to summation xi y i - n times xbar ybar upon summation xi square - n times xbar x square summation is i goes around 1 to n.

If I try to simplify, this quantity nothing i goes around 1 to n xi - xbar y i - ybar and this quantity in the denominator is summation xi - xbar whole square now keep in mind that this xbar and say ybar they are simply our sample mean, whatever the observations we had obtained based on the that I can find out there sample means, so xbar and ybar are known to us.

So now I can see one thing that when we stated our model y= beta0 + beta1x+ epsilon in that model this beta0 and beta1 were known to us, but now I can see that once I have got the observations using those observation I can find out the value of beta1. So this I take as an estimator of beta1 in simple words estimator means that the value of the parameters that can be obtain on the basis of given set of data.

So I have here parameters that is beta1, but is value is completely unknown to us now I am saying that using my observation I can compute the value of beta1 from this expression, which I have written here, so this is an estimator of beta1. So for the sake of simplicity let we try to rewrite here say beta1 hat = to sxy upon sxx where this sxy nothing but summation i goes from 1 to n xi - xbar y i - ybar and sxx is i goes from 1 to n xi - xbar.

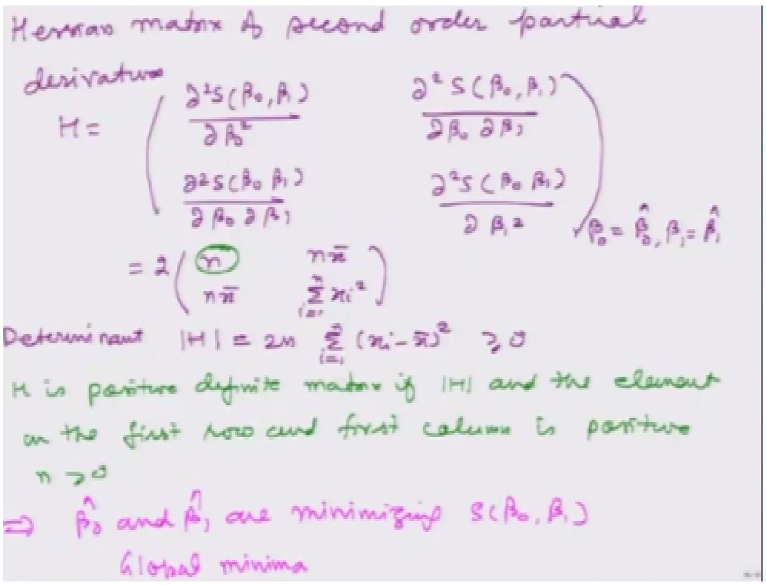
So now in the later lectures we are going to use this notation, so what I have seen now that we have obtain the value of beta1 has beta1 hat now the value of beta0 that we had obtained in the earlier slide here like this one this is going to be known to us only when beta1 is known to us or I try to write down here that this value of beta0 can be known to us if I try to replace my beta1 by beta1 hat like this.

So now using this expression I can again estimate my intercept term so this beta0 hat is an estimator of beta0. Both this beta0 hat and beta1 hat they have been obtained from the principles of least square or in this case we have a minimize the vertical distance between the observed values and the line something like here you can see we had minimize these thing.

So they are also known as direct regression estimators this beta0 hat is the direct regression estimator of beta0 and beta1 hat is the direct regression estimator of beta1 they are also called has least squares estimates or least squares estimators of beta0 and beta1. well we have obtain these thing, but we do not know whether the values of beta0 and beta1 that we have obtain as a beta0 hat and beta 1hat are they really minimizing my random errors are they are maximizing it.

So for that we have to find out the second order condition, here you can see I have got here two parameters and we are jointly estimating them.

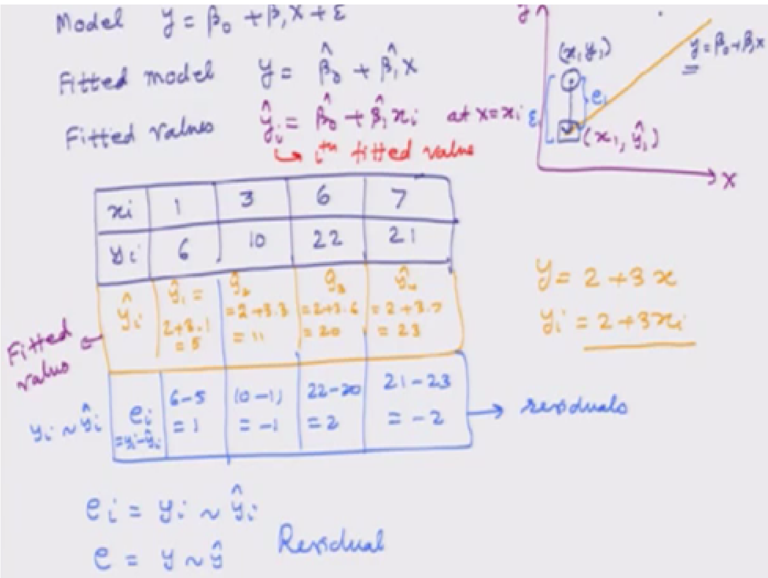
(Refer Slide Time: 17:39)



So we need to check about the bordered hessian matrix, so the hessian matrix of second order partial derivatives is defined here as H that can be a matrix of 2 by 2 with the partial derivatives of the second order with respect to beta0 and second order partial derivative of s with the respect to the beta0 and then beta1 and on the second diagonal the partial derivative of s with the respective beta1 square.

This matrix has to be obtained at beta0=beta zero hat and beta1 = beta 1 hat, so what I have to do I simply have to differentiate it again and then substitute beta0=beta0 hat and beta1=beta1 hat in the normal equation that we have to obtain here. In fact they are actually providing us a global minima, you can see we have obtain the value of beta0 and beta1.

(Refer Slide Time: 19:32)



So, if I try to write down compactly we had a model y=beta0+beta1x+epsilon and we have obtained a fitted model and this fitted model is y=beta0 hat, + beta1 hat x, and this model is called as a fitted regression model. Now after this I have to obtain the fitted values. What is this fitted value? You see when we conducted the experiment and we obtained the data.

Then there was some different between the observed data and that line, so if I try to make the earlier diagram here once again say this was your x and this was your y and this was your line and the observation they were lying somewhere here, here and so no. So if you try to observe this is suppose our x1, y1 and we expected that this value is going to lie somewhere here on this line y=beta0+beta1x.

So we had observed the values x1, y1 but I am expecting that this value should lie somewhere here. The value of y which is obtained using the observed value of xi this is the i'th fitted value. Well, let me try to explain simple example suppose I have got here a data, which I can write xi and yi, suppose I have here four sets of data I take suppose xi=1 and I obtain yi=6.

I take xi=3 and I obtain yi=10 I take xi=6x and I obtain yi=-22 and once I take xi=7 I obtain yi=21 Now suppose after fitting the model after obtaining the values of beta0 hat and beta1 hat on the basis of this four pairs of observation suppose I get here a model y=2+3x, so this means all this xi , yi they are also going to satisfy this model.

Now I can obtain here the value of yi hat, how I can obtain? for example my y one hat that is going to be, I will try to use this model. So this is going to be 2+3 xi, this is actually here 1, so this is going to be 5. Similarly if I try to obtain here y2 hat this is going to be 2+3 times here 3, so this is going to be 11, similarly y3 hat this is going to be 2+3 into 6 and this is 20.

And similarly y4 hat this is 2+3 into 7 this is going to be 23 After this I can write this point here has a x1 and say y1 hat, so these are nothing but my fitted values and if you try observe what are these values, I simply have fitted the model on the basis of given set of data then using the given values of xi I am trying to obtain the values of yi, which are yi hat, so yi hats are the values of y.

Which are obtained from the model and they are called as fitted values. So I can write down the fitted value here as y hat=beta0 hat+ beta1 hat and now I am using a given value at x=xi so this is my i'th fitted value. Let us try to see a different aspect, now I can find out the difference between yi and yi hat, so yi is the absolute value yi hat is the value of y, which is obtained from the model.

So in this case if you try see I try to be note by ei and suppose I define it yi - yi hat, so this becomes here 6-5 which is = here 1 this becomes here 10-11 Which is = -1 and similarly this becomes a 22-20 this is 2 and this becomes 21-23=-2 these values are called residuals.

I try denote it by ei, so ei is nothing but the difference between yi and yi hat, and in general I can define e as residual the difference between observed and fitted values so this is my residual, now this residual has a very important property. If you observe in this picture am saying the difference between y1 and y1 hat now as per this definition is nothing but e1.

Earlier we had denoted the same distance as epsilonone, so you can see here that this residues are going to act like as we have observed the random error in my data, remember one thing residuals are random variables, errors are random variables and am not estimating the random errors by residuals, but they will look like as if they are the values of random errors and these residuals helps us a lot in obtaining the information about my random errors and this will try to discuss in the forth coming lectures till then good bye.