Regression Analysis and Forecasting

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

Lecture 07-
Maximum Likelihood Estimation of Parameter in Simple Linear Regression Model

Welcome to lecture number seven, you may kindly to call that in the earlier lectures we had demonstrated the estimation of model parameters beta0, beta1 and sigma square using the principle of least-square and we add obtain the direct regression estimates of beta0, beta1 and sigma square. Now we use another approach, the maximum likelihood estimation.

And we demonstrate how to obtain the estimates of model parameters beta0, beta1 and sigma square, so we are going to talk about the maximum likelihood estimation of parameters.

(Refer Slide Time: 00:52)

Maximum likelihood estimation

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(\varepsilon_i - 0)^2}{\sigma^2}\right]$$

Likelihood function : $f(\varepsilon_1, \varepsilon_2 \cdots \varepsilon_n) = \prod_{i=1}^{n} f(\varepsilon_i)$

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{\varepsilon_i^2}{\sigma^2}\right]$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma^2}\right]$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right]$$

$$L^* = \ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

So you may kindly recall that we had considered the model yi=beta0+beta1xi+ epsilon i and we had assumed that epsilon i's are all iid identically and independently distributed following a normal distribution with mean 0 and variance sigma square, so what is a likelihood function? Likelihood function is nothing but the join probability density function of all the random variables.

So here we have the random variable epsilon and then with the joint probability density function of epsilon1, epsilon2, epsilon n will be the likelihood function of epsilon. We all know that the probability density function of epsilon i this is going to be 1 over sigma route 2 pie exponential of - 1 over 2 epsilon i- 0 whole square over sigma square.

Now when I try to write down the likelihood function, likelihood function is given by f of epsilon1, epsilon2, say epsilon n this is the joint probability density function of epsilon1, epsilon2 and up to epsilon n, since we assumed that all epsilon i's are independent so I can write down this joint probability density function as a product of i goes from 1 to n, f of epsilon i.

So let me note this likelihood function here as L and then the likelihood function can be written as 1 over product of i goes from 1 to n sigma route 2 pie exponential of - 1 over 2 epsilon i square upon sigma square. This becomes nothing but 1 over sigma square route of 2 pie, power of here n and exponential of- 1 over 2 summation i goes from 1 to n, epsilon i square upon sigma square.

The objective in the method maximum likelihood estimation is to obtain the values of the parameters in such a way such that this likelihood function is maximized here we want to estimate the 3 parameters beta0, beta1 and sigma square, but from this likelihood function you can see that still there is no appearance of beta0 and beta1. So, I tried to write down this likelihood function as an exponential of - 1 over 2 sigma square.

i goes from 1 to n, and I tried replace epsilon i by yi - beta 0 - beta1 xi whole square. Now the objective is very simple I want to maximize this L which is our likelihood function and I want to estimate the parameter beta0 beta1 and sigma square and since we know that lock function is a monotonic function that is a monotonic increasing function.

So you will see later on that it is easier to handle the log likelihood rather than likelihood function. So, at this moment itself I can take here a log of a likelihood, log here L and I am denoting it here by L star this is going to be - n by 2 natural log of 2 pie - n by 2 log of sigma square - on over 2 sigma square summation i goes log on to n yi – beta0 – beta1 xi whole square.

(Refer Slide Time: 05:38)

$L^* =$ function of β_0, β_1 & σ^2.

Normal equations

i) $$\frac{\partial L^*}{\partial \beta_0} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

ii) $$\frac{\partial L^*}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

iii) $$\frac{\partial L^*}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0$$

(i) $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$ or $n\bar{y} - n\beta_0 - \beta_1 n\bar{x} = 0$ $\quad \bar{x} = \frac{1}{n}\sum x_i$, $\bar{y} = \frac{1}{n}\sum y_i$

$\Rightarrow \tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x}$ * Same as

(ii) $$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$ * OLSE

(iii) $$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2$$ → different from

So, now you can see here that this L star which is a log likelihood, L star is a function of the parameters beta0, beta1 and sigma square. Now I try to obtain the value of beta0 beta1 and sigma square such that log likelihood is maximum so I use the principal of maxima and minima for which I find the normal equations like this, say I will simply partially differentiate the log likelihood functions with respective to beta0, beta1 and sigma square.

Once I try to differentiate it I get here - 1 over sigma square summation i goes from 1 to n, yi – beta0 – beta1 xi and here I get - 1 over sigma square summation i goes from 1 to n, yi – beta0 – beta1 xi times xi and here I get - n over 2 sigma square + 1 over 2 sigma squares, so this will become 1 upon 2 sigma power of 4 i goes around 1 to n yi – beta0 – beta1 xi whole square.

Next objective is that I will try to equate these normal equations to be 0 and I will try to obtain their solution, so I can do it at t his step, for example, so if I try to put it=0, substitute =2, substitute =0 then I try to solve it, so let me call this equation as equation number 1, this equation number second and this is equation number third.

So, the first equation if I try to solve it, this simply gives me summation i goes from 1 to n yi – beta0, - beta1 xi put it =0 and this can be written here as or n times y bar - n times beta0 - beta1 times, n times x bar and put it0 where this x bar is nothing but 1 over n summation n xi and y bar is 1 over n summation n yi.

Now once I try to solve this implies that the value of beta0 comes out to be y bar – beta1 x bar. I can treat this beta0 as the maximum likelihood estimator of beta0 provided beta1 is known, so far a while suppose that beta1 can be estimated by beta1 tilde, so now this beta0 tilde becomes the maximum likelihood estimator of beta0 well we will still have to verify whether this value will provide us minima or say maxima that we will try to do in the next slide.

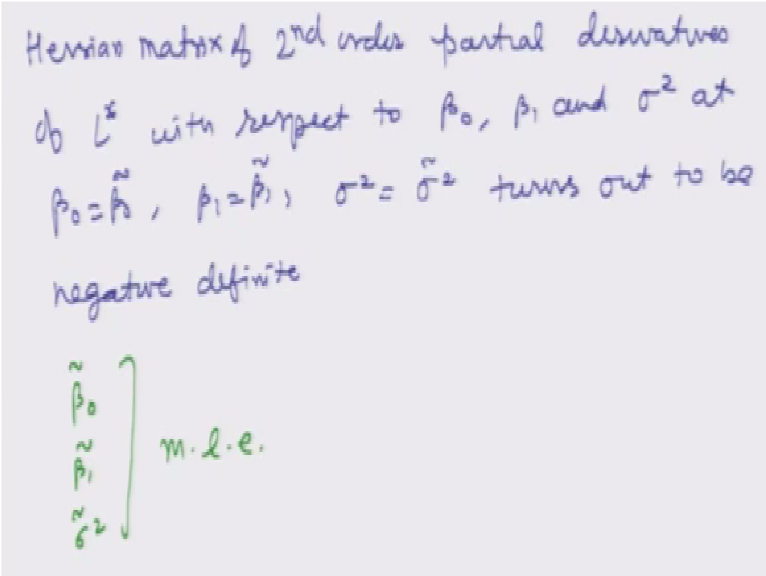
But here if you try to solve this equation number second, if you simply put it =0 exactly in the same away as we have done earlier this beta1 value will come out to be summation i goes around 1 to n xi - x bar yi - y bar upon summation i goes from 1 to n, xi - x bar whole square. So, I can call this beta1 as beta1 tilde and I will show you later on that this also the maximum likelihood estimator of beta1.

Now solving the third equation we get the value of sigma square as a 1 over n summation i goes from 1 to n yi – beta0 – beta1 xi whole square now this value is again going to be known only when beta0 and beta1 are known to us so I can replace beta0 by beta0 tilde and beta1 by beta1 tilde that we have operate above, this become the maximum likelihood estimator of the sigma square.

1 thing which I would like you to observe is that these thing, 1 is beta0 tilde and beta1 tilde they are same as ordinary least-square estimator of beta0 and beta1, if you recall earlier using the principle of least-square we had obtain the value of beta0 and beta0 hat and beta1 as beta1 hat and these 2 value are the same. But this is different form least-square estimator, the least-square estimator of sigma square.

If you recall that was 1 over n-2 summation i goes from 1 to n yi – beta0 hat – beta1 xi whole square, so the different between the maximum likelihood estimate and least- square estimate of sigma square comes out to be only because of the divisor here in this case the divisor is n and earlier the divisor was n-2.

(Refer Slide Time: 11:47)



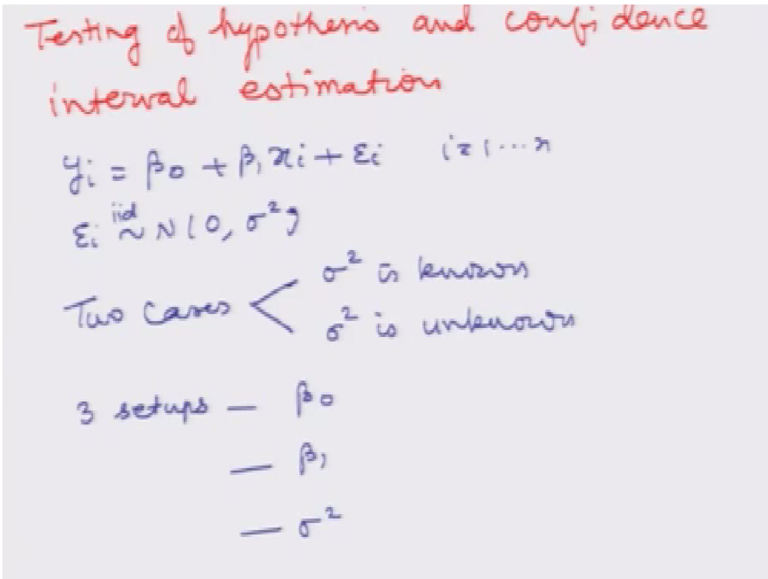
Now still we need to show that value beta0, beta1 and sigma square that we have obtained as beta0 tilde, beta1 tilde and sigma square tilde respectively they are maximizing the likelihood or minimizing the likelihood, so I would say simply that the hessian matrix of a second order partial derivatives of L star which is the log likelihood function with respect to beta0, beta1 and sigma square at beta0= beta0 tilde, beta1=beta1 tilde and sigma square = sigma square tilde turns out to be negative definite.

This exercise is almost the same as we have done in the case of least-square estimator, so this shows us that the value what we have obtained as a beta0 tilde beta1 tilde and sigma square tilde they are the maximum likelihood estimates of beta0, beta1 and sigma square. This is about the estimation part, now you can see using 2 differ methods we have obtain the estimates of the model parameters.

The idea of taking 2 different estimation procedure is as follows, we want to show that if you try to use any other estimation to procedure you may get a different estimator, it is only by chance here that the estimates of a intercept term beta0 and a slope parameter beta1 they turn out to be the same, but here also you have seen that the estimate of sigma square comes out to be different.

So similarly if you try to use any other estimation procedure there is a likelihood or there is possibility that you may get different estimates.

(Refer Slide Time: 14:29)



Now our next objective is how to conduct the test of hypothesis and confidence interval, so how we try to attempt on testing of hypothesis and confidence interval estimation. I assume here because the participant will have some favorable knowledge of test of hypotheses and confidence interval estimation and my objective is to demonstrate here how to conduct the test of hypothesis and confidence interval estimation over the model parameter say beta0, beta1 and sigma square.

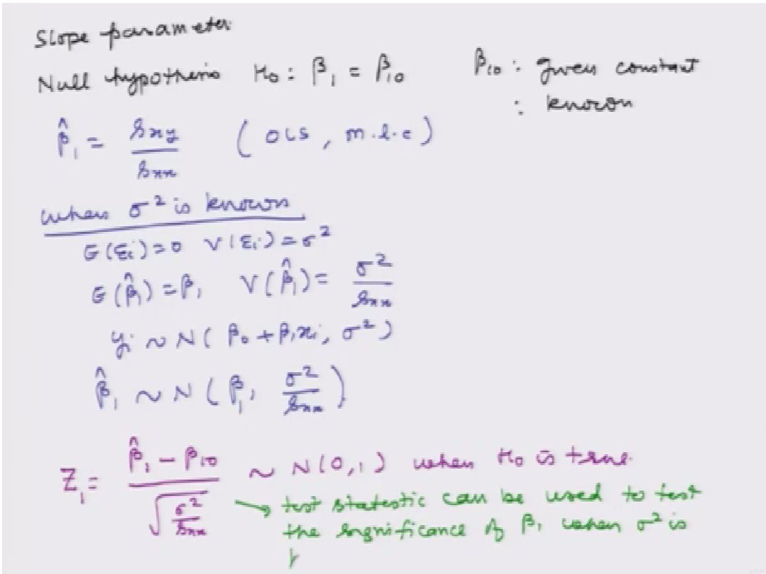
Just for the sake of understanding, our model is beta0+beta 1 xi+ epsilon i, and we have got n observations and we assume here that epsilon i is normal 0 sigma square and they are iids, they are identically and independent distributor. At this moment I would like to make here a simple comment, you have seen that when we are using the principle of least-square to estimate the unknown model parameters we do not require the knowledge of normal distribution.

When we attempted to estimate the parameter from the method of maximum likelihood then we require the assumption of a normality of random error components right from the first step. Now when we are going for testing of hypothesis and confidence interval estimation then we need the assumption of normal distribution right from the first step, now we will be developing the estimates of beta0 beta1 and sigma square.

So, all those tests are going to be based on this fundamental assumptions that normality assumption hold true. Now in this models setup I would say that we have here 2 types of possibilities 1 case is that we assume that sigma square is known to us and another possibility is that we assume that sigma square is unknown to us. Right, and then we will try to develop here a 3 setups.

1 for beta0 another for beta1 and another for sigma square, so we will try to develop the test of hypothesis as well as confidence interval for these 3 parameters under the 2 cases that when sigma square is known and when sigma square is unknown. So first of all I try to consider the case of a slope parameter and my null hypothesis is h not say beta1=beta1 0.

(Refer Slide Time: 17:22)



This beta10 is some given constant and this is assumed to be known. Before going forward I would like to inform you that this test of hypothesis and confidence interval estimation they play very important role in getting a good model, later on you will see that through testing of hypothesis and confidence interval estimation we try to judge whether a variable is important or not.

So this is a very important for us to understand that how we are going to develop the test and under what type of conditions we are going to use it. Now you can recall that you had estimated the slope parameter by sxy upon sxx and that was in both the cases whether it was ordinary least-square estimation or this was method of maximum likelihood estimation. Now I try to consider the first case that when sigma square is known.

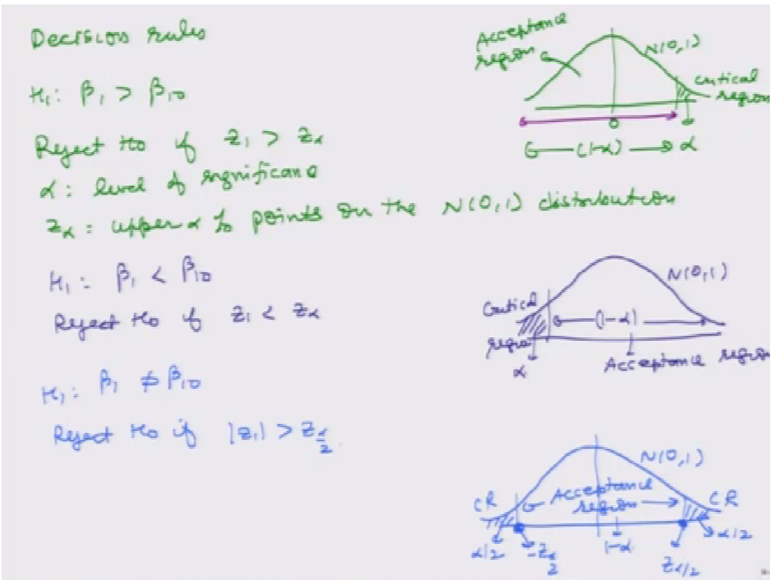
So under this thing since we have assumed that accepted value of epsilon i is 0 and variance of epsilon i is sigma square, based on that we have demonstrated that accepted value of beta1 hat is beta1 and variance of beta1 hat is sigma square upon sxx these are the expression that we had obtain earlier and we had also shown that beta1 hat is a linear function of y's and y's are following a normal distribution.

With mean beta0 + beta1 xi and means and variance sigma square. All these information that we already have established so now I can write down that since beta1 hat is also linear function of yi, so beta1 hat also follows a normal distribution with mean beta1 and variance sigma square upon sxx. So based on that now I can create here a statistics, now I can write down here.

That beta1 hat – beta1 0 divided by square root of sigma square over sxx that we know this is going to follow a normal distribution with mean 0 and variance 1, when h not is true. So let me denote this statistic as z1, so now z1 is the test statistics that can be used to test the significance of slope parameter when sigma square is known. So this test statistic can be used to test the significance of beta1 when sigma square is known.

Right how to get it done? We know that in the case of test of hypothesis if try to draw the distribution the entire distribution is divided into 2 parts.

(Refer Slide Time: 21:21)



1 is critical region and another is acceptance region, this is here mean the 0 and this is the curve of normal 0 1 so depending on the alternative hypothesis I can frame here different type of decision rules. Suppose my alternative hypothesis is beta1 greater than beta1 not, in this case the rule is reject h not, if z1 is greater than z alpha.

Where alpha is the level of significance. So this region here is actually alpha, and this region here is 1 - alpha, so we try to calculate the value of z1, and we try to see if the value of z1 is going to lie in this region we accept it and if it is going to lie in the critical region than we reject it, and here this z alpha is the upper alpha % points on the normal 0 1 distribution.

Next I try to take another case in case if my h 1 is beta1 less than beta1 not, so in this case this normal0 1 curve will like this, but the critical region is going to be on left hand side, so this is our critical region and the size of critical region that we are going to consider this is alpha, so this is your acceptance region of size 1 - alpha, acceptance region and in this case I would say that reject h not if should not if z 1 is less than z alpha.

The third case will be that we try to consider here h1 is see here beta1 not equal to beta1 not. In this case the normal 0 1 curve will have the critical region on both the sides, this is our critical region and this our acceptance region, and this is of size alpha by 2, this is of size alpha by 2, and this is of size 1 - alpha. So these are the points which are - z alpha by 2 and this is the point which is z alpha by 2 because the mean lies over here.

So in this case I would say that reject h not if z1 mode of a z1 is greater than z alpha by 2. So this is how we are going to take a decision. We stop here and in this lecture, I have tried to explain you in detail that how we going to construct the test of hypothesis. We have considered the case for the test of hypothesis h not beta1=beta1 not and I have given you the development of the statistics.

And how to take the decision on 3 types of alternative hypothesis. Now I will try to take some more cases for example when sigma square is unknown and then the test hypothesis based on intercept term and sigma square in the next lecture, till then good bye.