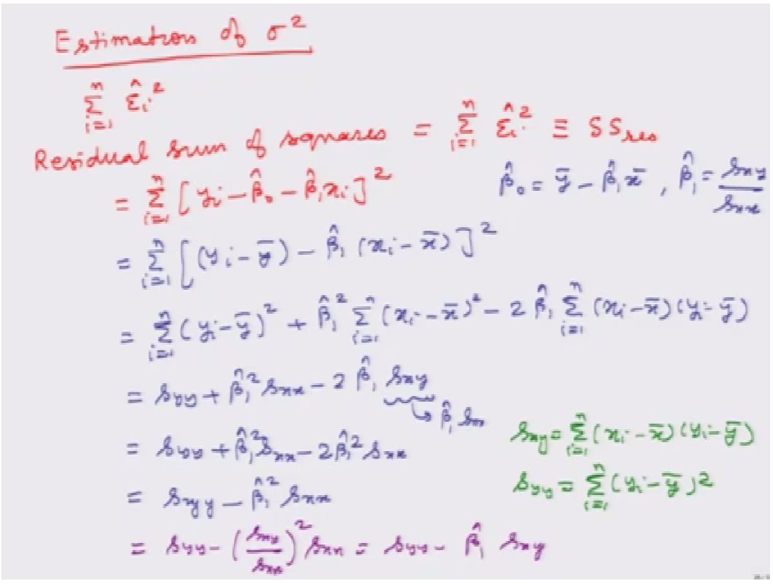
Regression Analysis and Forecasting
Prof. Shalabh
Department of Mathematics and statistics
Indian Institute of Technology – Kanpur

Lecture – 06
Estimation of Parameters In Simple Linear Regression Model (continued)

Welcome to the lecture number six if you recall on the last lecture we had estimated the parameters beta0 and beta1 we had to further investigated their statistical properties we had proved that beta0 hat and beta1 hat are the unbiased estimators of their respective parameters and we also found their variances.

In the model they were three parameters beta0 beta1 and sigma square out of those three parameters we have estimated 2 parameters but the third parameter is still left.

(Refer Slide Time: 01:02)



So we start with the estimation of third parameter which is sigma square, there is a difference in estimating sigma square and the 2 parameter beta0 and beta1 by using the principal of least squares there is no involvement of sigma square in the function s beta0 beta1 that we had defined earlier, so it is not possible to differentiate s with respect to sigma s square, put it =0 and solve the equation to get the value of sigma square.

In order to obtain the lease s square estimate of sigma square we start with some estimator based on our guess, now the next question is how are we going to make a guess, we had discussed that residuals are the difference between observed and fitted values and sigma square is the variance of epsilon i's so 1 option is that I can consider quantity like summation i goes around 1 to n, epsilon i hat square and then

We try to take this expectation this will come out to be a as the function of sigma square, then we adjusted to obtain a good estimator of sigma square. Before doing that let me define it, so we try to define something called residual some of square and that is defined as some of square of the residual, so call it as ss res, ss res means some of square due to residual RES means it is a short form of residual, but before going further

Let us try to understand it this quantity have some important role in the regression analysis so we would like solve it further and let us and obtain different possible forms of some its square due to residuals. If you try to see here this quantity is nothing, but i goes from 1 to n see here yi - beta0 hat - beta1 hat xi whole square. So, I can now make it like this, I try to substitute the value of beta0 hat.

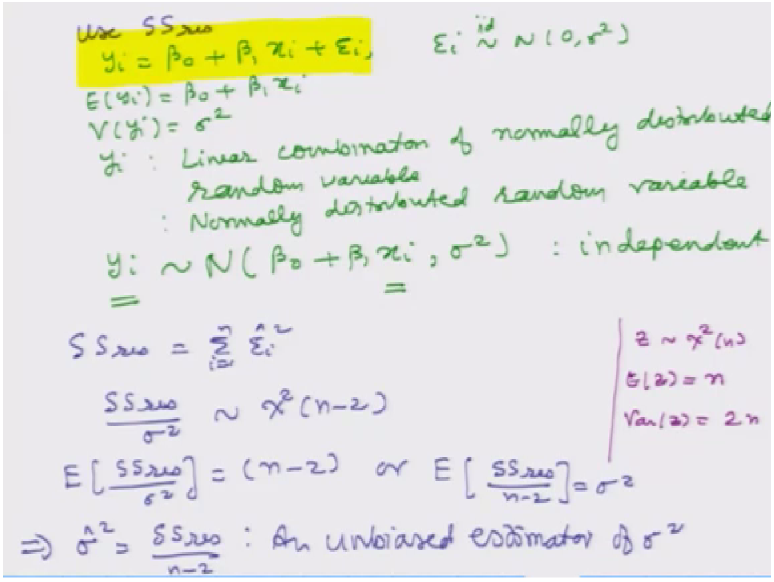
Which =ybar - beta1 hat xbar and then this becomes y1 - ybar - beta1 hat xi - xbar whole square and this can be further written as summation i goes from 1 to n yi - ybar whole square + beta1 hat square summation i goes from 1 to n xi - xbar whole square - twice of beta1 hat summation i goes from 1 to n xi - xbar yi - ybar.

If you recall earlier we had defined the quantity like sxy this sxy was defined here has a summation of i goes from 1 to n, xi - xbar yi - ybar on the same lines I can define syy as i goes from 1 to n, yi - ybar whole square, so this quantity now becomes syy + beta1 hat square and this quantity is in nothing but your sxx - twice of beta1 hat sxy and we had seen that beta0 hat is this and beta1 hat is sxy over sxx.

So can write down this sxy as a beta1 hat times sxx, so this becomes here syy + beta1 hat square sxx - twice of beta1 hat square sxx, so this can be written as syy - beta1 hat square sxx, this is 1 possible form, now I can replace here the value beta1 hat is = sxy upon sxx and this I can further solve as syy – xxy over sxx whole square sxx and this will come to syy - beta1 hat sxy.

So this residual some of the squares has got several popular forms and depending on the need and situation we try to use them suitably. Now our next question is how to obtain the estimate of sigma square?

(Refer Slide Time: 07:05)



So let us try to use some of square due to residual to find out an estimate of sigma square but before that we need to find out the distribution of yi so since we assume that our model is beta0 yi= beta0 = beta1 xi + epsilon i and we have assumed that epsilon i are following a normal 0 sigma square distribution and they are IID'S. I can write down here that expected value yi is nothing but beta0 + beta1 xi.

And we also have earlier that variance of yi = sigma square more over yi's are the linear combination of normal random variable of normal or normally distributed random variable which in our case is epsilon, so I can also write that yi will also be normally distributed random variable, I can write down here yi's are going to a fallow a normal distribution with mean beta 0 plus beta1 xi and variant sigma square.

Now are they also IID, if you try to see they are independent but they are not identically distributed why because the distribution of yi is depending on xi so as the observation changes the mean of the normal distribution also changes, so yi's are independently distributed following our normal distribution with mean beta0 + beta1 xi and variant sigma square.

Once I can write this thing then we also know that some of squares of normal random variable follow a chi-square distribution. So if there are n random variables each of them is following a normal random variable then there are some of square will follow a chi-square distribution with n degrees of freedom, so incase if I try use this thing, this some of squares due to residual.

If you try to see this is nothing, but the sum of squares of normal random variables, so I can write down here that ss residual divided by sigma square this will follow a chi-square with n - 2 degrees of freedom. one question comes how does 2 comes into picture, since I know that epsilon i has they are depending on 2 unknown parameters, beta0 and beta1, which you are estimating as beta0 hat a beta1 hat.

So there are 2 unknown parameter which are being estimated in finding out the epsilon I hat square, so that is why there are 2 constraints and so that degrees of freedom are reduce by 2 and the degrees of freedom of this chi-square random variable are n - 2. Just for the sake of your information and an quick review means we know that if there is some random variable z which following a chi-square distribution with n degrees of freedom then its mean that is respected value of z is n, that is the number of degrees of freedom.

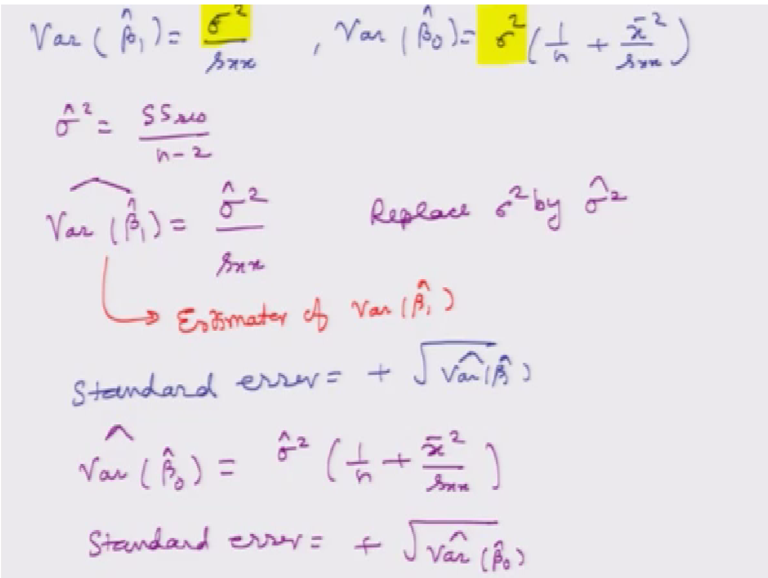
And variance of z it is nothing but twice of n, that is the twice of the number of degrees of freedom. So if I try to find out here the mean of this quantity so I can write down here expected value of ss res divided by sigma square this is following a chi-square distribution with and - 2 degrees of freedom so its mean is going to n - 2. I can write down here or I can write down here expected value of ss res, divided by n – 2.

Is = sigma square, so this implies that sigma square hat = ss res, divided by n - 2, so this is an unbiased estimator of sigma square. If you try to observe what we d1 we started with the model yi= beta0+ beta1 xi + epsilon i and based on that we have three parameter beta0, beta1, and sigma square all of them were unknown to us what we did we collected a sample of data and we had an observation something like x1, y1, x2, y2, xn, yn.

Based on that we have obtained the estimate of beta0, estimate of beta1 and now we have also obtained the estimate of sigma square these three parameters can be estimated using these three estimators, so now we can say that my model is known to us. There is now another question when we are trying to estimate my model parameters for example we have estimated beta0 by beta0 hat and beta1 by beta1 hat.

How I can say that these are good or bad we have established that they are unbiased estimator and we also have found there variances, and we had obtained there variances like this variance of beta1 hat this was observed as sigma square upon ssx and variance of beta0 hat was observed as sigma square, 1 over n + xbar square over sxx.

(Refer Slide Time: 14:05)



Once we are trying to estimate the parameters on the on the basis of a sample I would also like to know what is the variability of my estimators because there can be more than 1 estimator for the same parameters that can be unbiased, but then how to choose among them. I would try to choose the estimator which has got a small variance so now I have got these 2 variances, but these are the values of the variance in the entire population.

If you try to see here this is depending on sigma square here and see here and sigma square is the value in the population, so these 2 variances are going to be unknown to us so my next objective is how do I estimate the standard errors of beta1 hat and beta0 hat. 1 option is that now since we have obtained that sigma square hat is ss res divided by n - 2, so 1 option is that I can write down the variance of beta1 hat as a sigma square

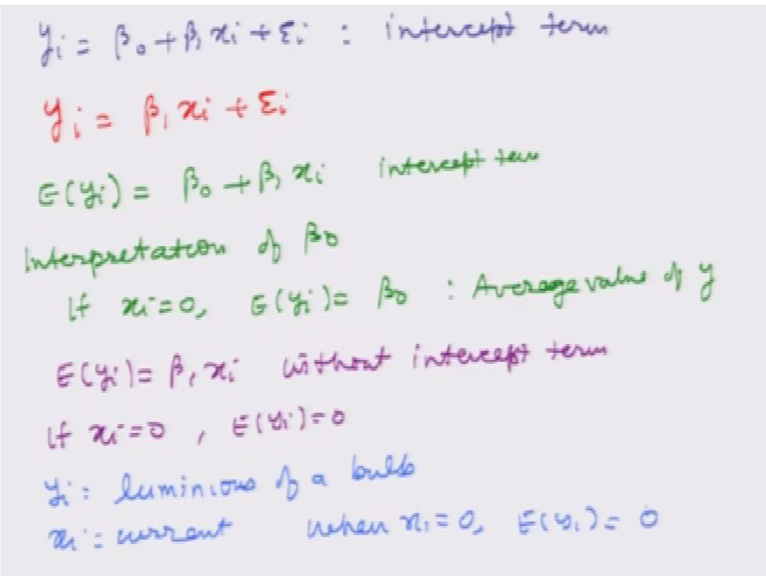
Upon sxx, but in this case I don't know the this variance, so I can do 1 thing that I can replace sigma square by sigma square hat, so this gives me an estimator of variance of beta hat, so now this is an estimator of variance of beta1 hat. Now what is the advantage that once I get a set of data I can estimate my model parameter beta1 using the ordinary least square estimator sxy upon sxx and now I can also provide its standard error.

Just by taking the positive square root of estimate of variance of beta1 hat, so this will also give us an idea that what is the performance of my estimator in terms of the variability. Similarly I can also find out the estimate of the variance of beta0 hat, what I have to do I simply have to write down the variance of beta0 hat and I have to replace sigma square by sigma square hat and if I want to find out its standard error it is very simple just try to take the positive square root of estimate of variance of beta0 hat that is all.

So now I have got a both the estimators beta0 hat beta1 hat I have shown that they are unbiased and I also have found their standard errors. So once I try write down a model give its parameter find out the estimate of all the parameter as well as standard errors up to certain extent we have obtained a model. After this thing there are some other considerations like test of hypothesis confidence interval estimation on the estimated parameters.

So that we will try to continue with those things, but before that let me try to explain you 1 more thing, we have considered here the simple linear regression model.

(Refer Slide Time: 18:34)



Yi= beta0 + beta1 xi + epsilon I, right this is a model which has got interceptor. Now there can be another situation where we need to consider a model without an interceptor, in that what would happen that I would try to consider the model something like yi=beta1 xi +

epsilon i the first question come, what is the difference between the 2? Why should I consider the intercept term some time to be present or some time to be absent?

I will try to take up simple example and I would try to explain you for example when I consider an example of yield of a crop, for example yield of a crop depends on the quantity of the fertilizer, rain fall, temperature so on in this case if we try to put some fertilizer in the soil means our crop will increase, but on the other hand I do not put any fertilizer the soil itself has some inherent fertility and because of this I will get some yield, if you try see here what is the interpretation of here yi.

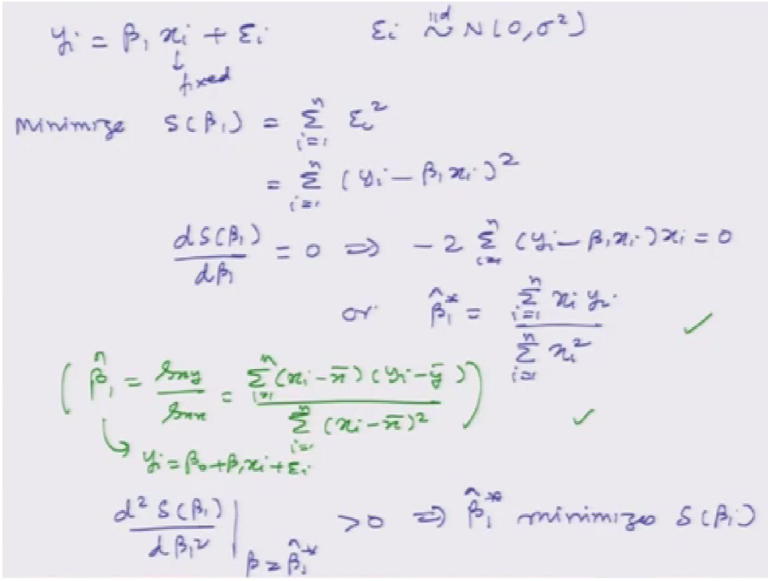
If I try to write down here the model in terms of expectation, if I have the intercept term in the model this model is like this with intercept term and this model is something like beta1 xi if I do not have the intercept term, so if you try to see what is the interpretation of beta0, beta 0 is nothing, but if I say if xi= 0 then expected value of yi= beta0, so that mean beta0 is nothing but the average value of y.

In this case when am trying to put 0 fertilizer in the soil am I getting 0 outcome, on the other hand if I try take the model without intercept term that something like beta1 xi without intercept term, then if xi= 0 then expected value of yi= 0.So if I try to take an example to explain the 2 situation I can say in the case of crop when I put no fertilizer in the soil.

Still I get something so my outcome is not 0, so the average yield is not going to be 0, so in that case I would like to have model with an intercept term. On the other hand if I say that y is my here something like the luminous of a bulb and xi is my current we all know that when we witch on the switch, then some current goes into the bulb and then the bulb glows, but if I put no current inside the bulb the bulb will not glow and the luminous will 0.

So in this case I know when I put xi is equal to 0 then my average value of yi is always 0, so in this case I would like to have a model without an intercept term. So there are situations in practice where some time I have to consider a model with intercept term and some time without intercept term and based on that I would like to show you in this simple frame work that what happens to the estimate of parameters.

(Refer Slide Time: 23:15)



So now in case if I try to consider here a model yi= beta1 xi= epsilon i all my assumption remain the same as earlier that beta1 is a parameter xi is my face it is non-scholastic and epsilon i are my IID normal 0 sigma square random errors. Right in this case my objective is to find out the parameter value beta1, so I again use the principal of least square and I try to minimize the sum of a squares which I try to denote.

Now here s of beta1 i goes from 1 to n epsilon i square and this is nothing but your i goes around 1 to n yi - beta1 xi whole square. When I try to differentiate it with respective beta1 and put it = 0 this gives me - twice of summation i goes around 1 to n yi - beta1 xi, say xi = 0 and when I try to solve it this gives me the values of here beta1 has a summation i goes from 1 to n xi yi upon i goes from 1 to n xi square.

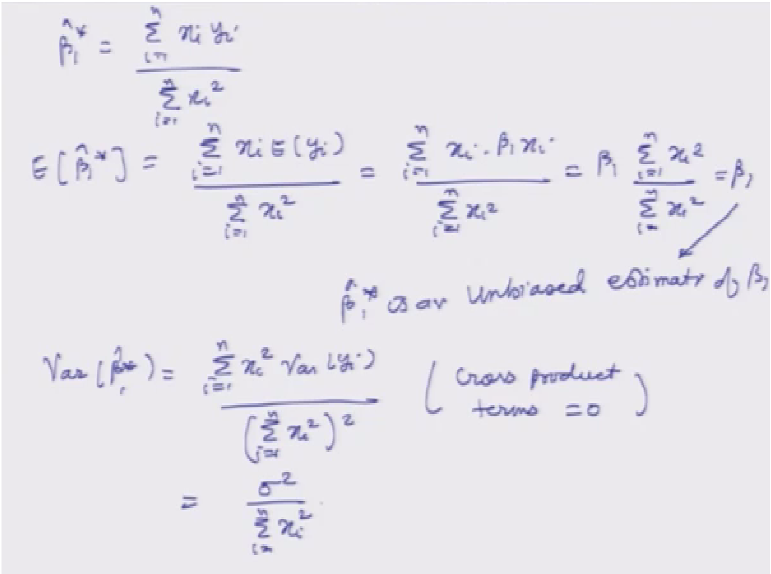
This an estimator of beta1 so I can denote it as a beta1 say star hat I am trying to use here a notation star just to distinguish beta1 hat from the beta1 hat that was obtained in a model with intercept term. Because for the sake of clarity I can write down here that your beta1 hat is sxy upon sxx which was summation xi - bar yi - ybar upon summation xi - xbar whole square.

And this was obtained in the model yi= beta0+beta1 xi + epsilon I, so that there is no confusion between the 2 symbols of the beta1 hat. So now what do you observe here does this expression and this expression are the same not really they are very different, so the model of the story is that when you are trying to fit a model with intercept term and model without intercept term.

There is one parameter that the slope parameter beta1 is common, but their estimators in the 2 models they are different that you have to keep in mind. Now it is also not difficult to show that the second order of derivative of beta1 with respective beta1 square at beta= beta hat star this comes out to be greater than 0 so this implies that beta1 hat star minimizes as beta1.

Now this beta1 star is the ordinary least square estimator of beta1 when we have no intercept in the model.

(Refer Slide Time: 26:54)



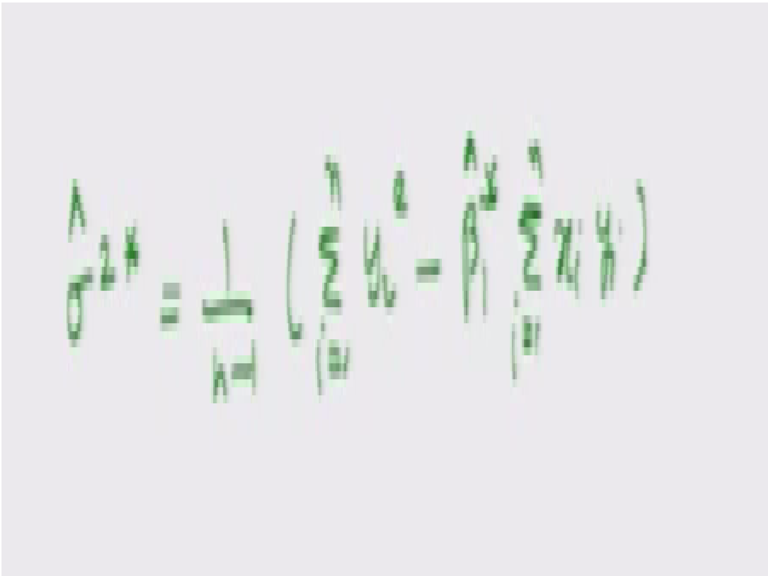
So now at the same time we can do a similar exercise and we try to establish the statistical properties that is finding out the unbiasedness character and the variance of beta1 star, so here is now here summation i goes from 1 to n, xi yi upon summation i goes around 1 to n xi square. So I can find out expected value of beta1 hat star this comes out to be summation I goes around 1 to n xi expected value of yi upon summation.

I goes from 1 to n xi square, and this quantity is nothing but i goes from 1 to n, xi an expected value of yi is nothing but beta1 xi, upon summation i goes from 1 to n, xi square. This quantity is nothing but summation i goes from 1 to n xi square upon summation i goes from 1 to n xi square and this is same as beta so this remains unbiased estimator, so beta1 hat star is an unbiased estimator of beta1.

That is established so there is no change in the property of unbiasedness of the slope parameter of the 2 estimators in the case of model with intercept term and model without intercept term. Similarly if I try to find out here the variance of beta1 hat star this is nothing but summation i goes from 1 to n xi square variance of yi divided by summation i goes from here 1 to n, xi square plus whole square.

And the gross product terms becomes equal to 0 because we have assumed that y 1, y2, yn are independent and we already established that variance of yi is sigma square so once I substitute this thing this comes out to be nothing but sigma square I goes from 1 to n summation xi square, and in this case if you also want to find out the estimate of sigma square that can be also obtained quite easily sigma square and let us try to say star.

(Refer Slide Time: 29:24)



So this can be obtained here as 1 upon n - 1 summation i goes from 1 to n yi square - beta1 hat say star summation I goes from 1 to n xi yi. So you can see here that even the estimator of sigma square in case of model without intercept term is quite different than the form of the estimator of sigma square when there is an intercept term in the model.

So, the model of the story is that if you do not want to consider the intercept term in the model please don’t do like this that you fit a model with intercept term and you simply substitute the intercept term =0 and use the same estimates. You will need to find the estimates of the parameter separately and the standard errors of the parameter estimates will also change.

So we stop here in this lecture and in the next lecture we will try to consider the maximum likelihood estimation of the parameters beta0, beta1 and sigma square and we will try to see how does it make a difference when we try to use different estimation techniques to estimate the parameter thank you and till then good bye.