Essentials of Data Science with R Software - 2
Sampling Theory and Linear Regression Analysis
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Linear Regression Analysis
Lecture - 49
Multiple Linear Regression Analysis
Test of Hypothesis and Confidence Interval Estimation on Individual Regression Coefficients

Hello friends, welcome to the course Essentials of Data Science with R software 2, where we are trying to learn the topics of Sampling Theory and Linear Regression Analysis.

In this module on Linear Regression Analysis, we are going to continue with our chapter on the Multiple Linear Regression Analysis. So you can recall that in the earlier lecture, we had talked about the estimation of the parameters and we had seen that how those concepts can be implemented in the R software.

So, now we are at a stage where we have estimated the parameters the regression coefficients as well as the variance. Now, the next step is the test of hypothesis and confidence interval. So, now in multiple linear regression model, test of hypothesis plays a very important role in making different types of very important conclusion about the model.

For example, in case if you have taken some number of variables or you have selected some variables in your model, you would always like to know whether those variables are important or not, or the data which we have collected on those variable is it really helping you in explaining the variation in the value of response variable.

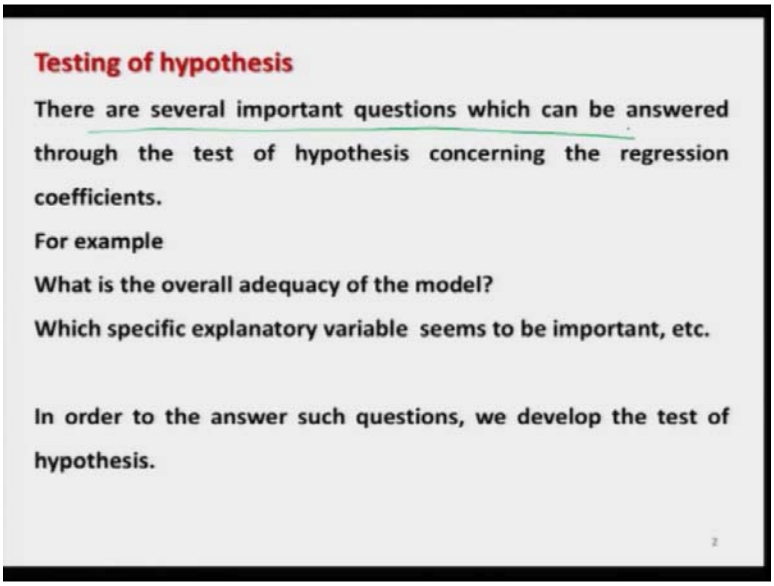
So, basically, you would like to retain only the important variables which are contributing in the model in some way. How to get it done, that is the question now, and this is what we are going to do with the test of hypothesis.

So, I will be considering two types of test of hypothesis; one is the test of hypothesis on a single regression coefficient, and I will be talking of the test of hypothesis when we have more than one regression coefficients. Now, the next question is, why not I am saying that I am going to consider here the test of hypothesis for the σ^2?

Well, that we already have covered. And whatever we have done in the case of simple linear regression modeling about the test of hypothesis and confidence interval for σ^2, the same story continues here.

The procedure is the same, concept is the same, command is the same, package is the same. So that is why, now I am going to consider about the test of hypothesis and confidence interval estimation only for the regression coefficients. Well, so let us begin our lecture.

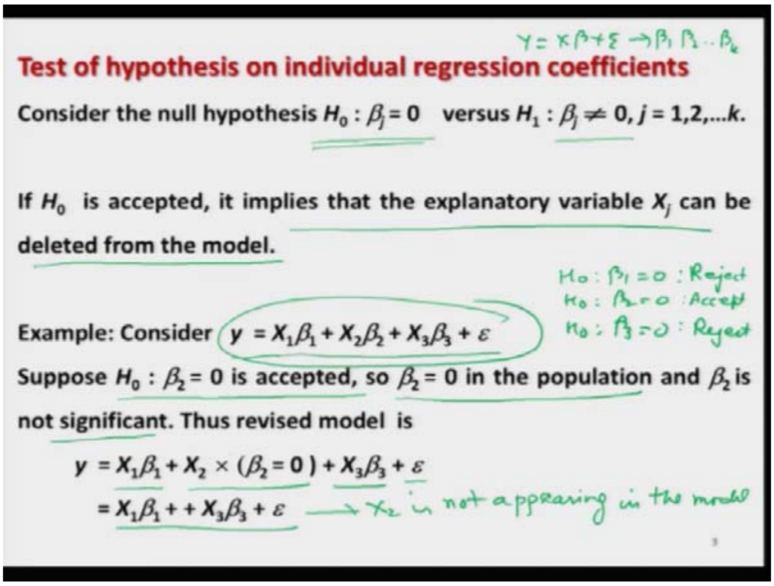
(Refer Slide Time: 02:53)



So, there are several important questions which can be answered through the test of hypothesis and they are concerning the regression coefficients.

For example, if we want to know what is the overall adequacy of the model or we want to know which specific explanatory variable seems to be important and similarly, there are different types of questions which can be answered. So, in order to answer such question, we would like to develop the test of hypothesis for the regression parameters, right.

(Refer Slide Time: 03:27)



So, first; we are going to consider the test of hypothesis on individual regression coefficients. Actually, we are going to consider two types of test of hypothesis; one for the individual regression coefficients and another will be analysis of variance, where we try to test the equality of all the regression coefficients, right.

So, in this case, we try to consider the null hypothesis $H_0 : \beta_j = 0$ versus the alternative $H_1 : \beta_j \neq 0$, and you can remember that you had the model $y=X\beta + \varepsilon$, which had the parameter $\beta_1, \beta_2, \ldots$ up to here β_k.

So, we are trying to test any of this $\beta_1, \beta_2, \ldots, \beta_k$ here. And, what is the interpretation of the acceptance of hypothesis? If, H_0 is accepted then it implies that the explanatory variable X_i which is corresponding to β_j can be removed from the model.

What does this mean? For example, suppose I have a multiple linear regression model with three variables X_1, X_2 and X_3 which I can write like this $y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon$. And now, suppose I try to test here say three hypothesis, $H_0 : \beta_1 = 0$ $H_0 : \beta_2 =0$ and H_0 $\beta_3 =0$.

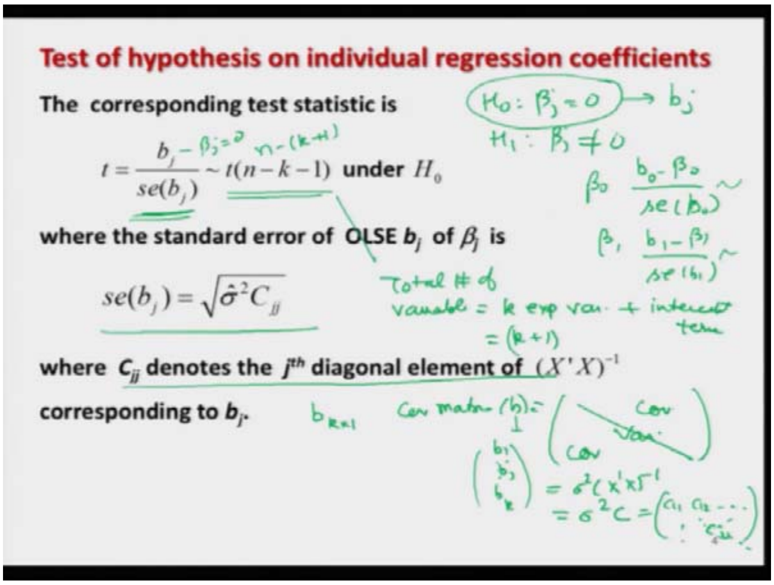
And suppose, the first hypothesis $H_0 : \beta_1 =0$ is rejected, $H_0 : \beta_2 =0$ is accepted, and $H_0 : \beta_3 =0$ is rejected. Now, what will happen? Once I say that $H_0 : \beta_2 =0$ is accepted; that means β_2 is almost 0 in the population. That means, β_2 is not significant.

Well, what does this mean? Now, if you try to put the same value in the model here in this model, now this can be written as y = X_1 β_1 + X_2 x (β_2 =0), which is coming from the test of hypothesis + β_3 X_3 + ε. So finally, this comes out to be X_1 β_1 + X_3 β_3 + ε. What does this mean?

Now you can see here, that the variable X_2 is not appearing in the model. Means, you can imagine that since β_2 is close to 0; that means, the rate of change in the average value of y with respect to X_2 is very small and possibly ignorable, and that is why we can believe or we can interpret that X_2 is not an important variable.

Hence, this variable can be removed from the model and the revised model will have only X_1 and X_3 . So, this is how you can see that the test of hypothesis plays an important role in the multiple linear regression analysis. It helps us in identifying the important variable in the model or it helps us in the selection of important variables in the model, ok.

(Refer Slide Time: 06:38)



So, now let us try to construct the test of hypothesis. So, we are going to test here as we discussed, H_0 : β_j = 0 versus H_1 : β_j ≠ 0, right. So, now you can recall that we already had constructed this statistics in the case of simple linear regression model. And, there if you remember, we had taken for example, (b_0 - β_0) /se(b_0).

So, this was the test statistics for β_0, and similarly for β_1 it was $(b_1 - \beta_1)$ se(b_1), where b_0 and b_1 are the ordinary least square estimator of the intercept term and slope parameter in a simple linear regression model, and depending on whether σ^2 is known, σ^2 is unknown, we had used the z statistics or t statistics.

So, here also we have the same thing. So, once we are trying to test the hypothesis H_0 : $\beta_j = 0$, where β_j has been estimated by OLSE or MLSE as b_j, right. So, the statistic which can be used here is t statistics. Why? Because, you have got only the sample of data and nobody is going to explain you or inform you what is the value of σ^2.

So, you need to estimate the value of σ^2 from the sample itself,, right. So, the t statistics can be framed as $b_j - \beta_j = 0$, right. So, this becomes b_j /se(b_j). And, this will follow here a t distribution with n - k - 1 degrees of freedom, right. Why there is n - k - 1 degrees of freedom?

Because, you are trying to include an intercept term in the model. So, your total number of variables of variables are say k explanatory variables + intercept term. So, that is why the number here is k + 1, right. And that is why we have written this thing as n - k + 1, like this. And then, how to obtain the standard error of b_j ?

So, you can remember that we had obtained the ordinary least square estimator b, which was a k x 1 vector and we had obtained the covariance matrix of b, right, where we had discussed that the diagonal elements are going to indicate the variances and off diagonal elements are going to indicate the covariances.

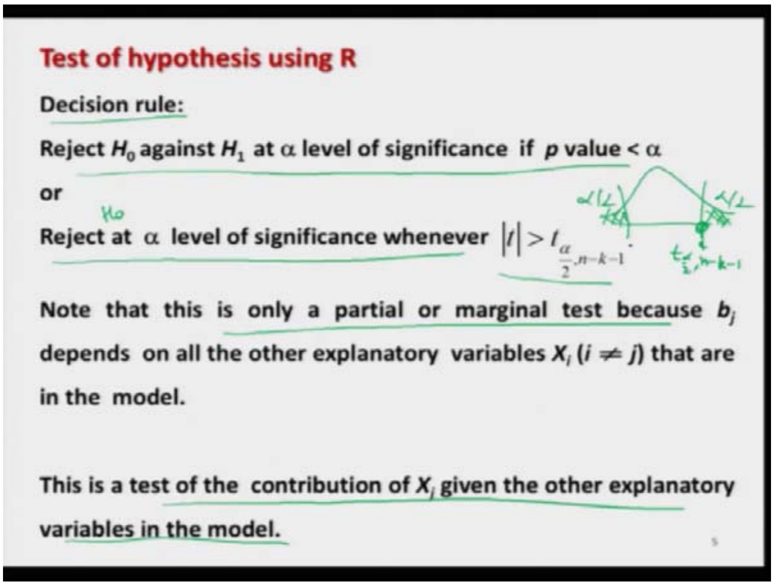
So, now if you try to take here b, b is a vector like b_1 , b_2 ,…, b_k and somewhere it will be here b_j . So, whatever is the diagonal element of this covariance matrix of b that is going to inform us the variance of b_j .

And if you try to take the positive square root, you will get the standard error of b_j . And you can remember that this covariance matrix of b was $\sigma^2 (X'X)^{-1}$, right. So, if I try to write down this matrix here I say, suppose if I try to write down here C matrix.

So, C $= (X'X)^{-1}$. So, now this C matrix can be written as C_{11} C_{12} up to here like this, and somewhere on the j diagonal the element will be C_{jj} . So, I can write down very simply that the $se(b_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$, where C_{jj} is the j diagonal elements of the matrix $(X'X)^{-1}$, right.

So, this is now my test statistics for testing the significance of $H_0 : \beta_j = 0$.

(Refer Slide Time: 10:48)



Now, how to find the decision rule? So, now, we have two approaches, in that first approach is when we are trying to use the software, so in that case the software will give us the p value, that we already have discussed what is p value and how to take a conclusion. So, now I will not repeat it again, but I will say the simple decision rule is reject H_0 against H_1 at α level of significance if p value is smaller than α.

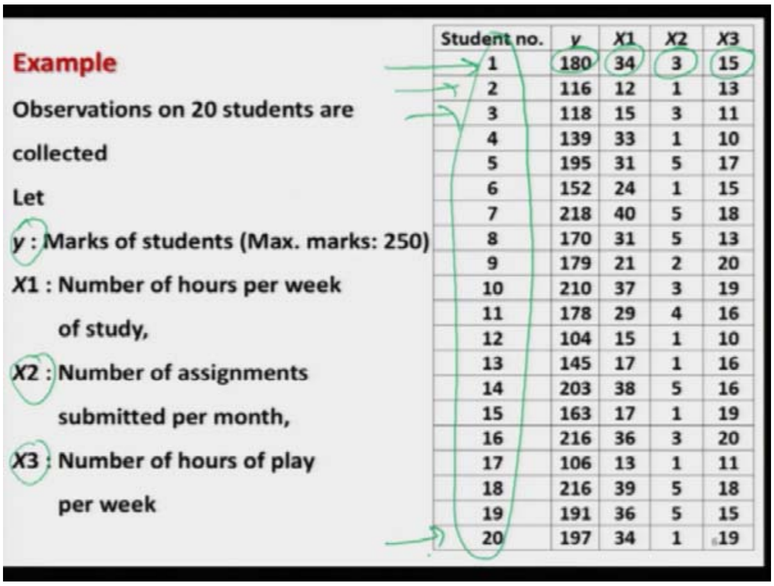
Or, if you come through the classical statistics and you try to divide the region into two parts; acceptance regions and rejection region and you try to take the type one error as α, so that the region of rejection is on both sides of the t distribution, both sides are having $\alpha/2$ area. So, in this case, I can say the H_0 is rejected if the calculated value of the statistic lies in the region of rejection, either here in the shaded area, right.

So, if the critical value which is obtained from the probabilities of t table, it is here given like this, $t_{\alpha/2}$ at n - k - 1 degrees of freedom. Then I can say that reject H_0 at α level of significance whenever $|t| > t_{\frac{\alpha}{2}, n-k-1}$.. So, essentially I am trying to say that reject H_0, right, ok.

So, if you try to see this test, so you can observe that this is only a partial or marginal test. Why? Because, b_j is not independent; means the j th regression coefficient that you have used here and you have used its standard error this is not an independent value of parameters, right, but it depends on all other parameters and all other variables also, right. So I can consider or we can consider this test as a contribution of X_j given the other explanatory variables in the model.

So, we have here information on X_1 , X_2 , …,X_k and then we have parameters say β_1 , β_2 ,…, β_k , but we are trying to take out only one parameter β_j , and out of this complete information we are trying to consider only one regression coefficient b_j , and then we are trying to construct the test of hypothesis for that, ok.

(Refer Slide Time: 13:02)



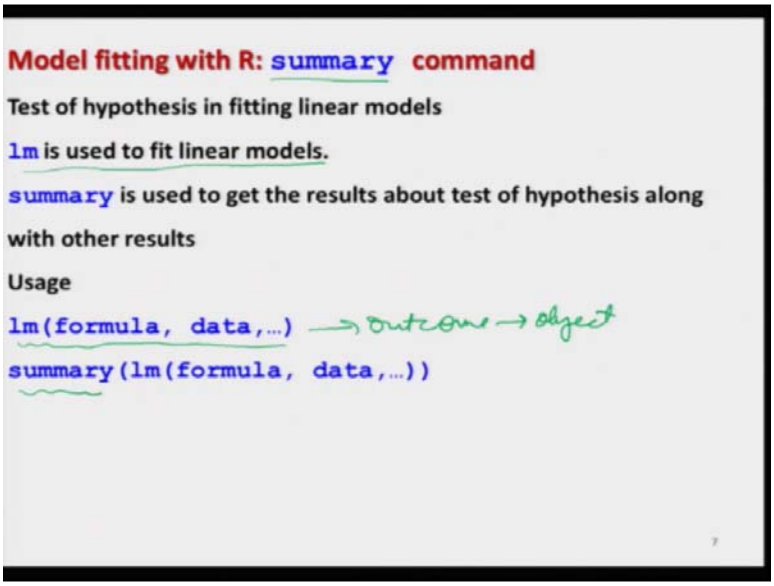
Student no.	y	X1	X2	X3
1	180	34	3	15
2	116	12	1	13
3	118	15	3	11
4	139	33	1	10
5	195	31	5	17
6	152	24	1	15
7	218	40	5	18
8	170	31	5	13
9	179	21	2	20
10	210	37	3	19
11	178	29	4	16
12	104	15	1	10
13	145	17	1	16
14	203	38	5	16
15	163	17	1	19
16	216	36	3	20
17	106	13	1	11
18	216	39	5	18
19	191	36	5	15
20	197	34	1	19

So, now I try to do one thing, I try to take one simple example and would try to show you that how you can conduct this test of hypothesis in the R software.

So, I am going to take here again the same example that we have considered in the last couple of lectures that we have data of students, there are 20 students, and we have collected the data on their marks obtained in an examination out of 250. Then the number of hours per week of study which is denoted as X_1 , and number of assignments which they have submitted per month that is X_2 , and the number of hours the student has played per week that is denoted by X_3 .

So, if you try to take a student, number 1, this means student number 1 has got 100 180 marks out of 250 and the student has studied 34 hours in a week, the student has submitted 3 assignments per month, and the student has played for 15 hours in a week. And similarly, the similar data is for the 20 students here, right, ok.

(Refer Slide Time: 14:07)

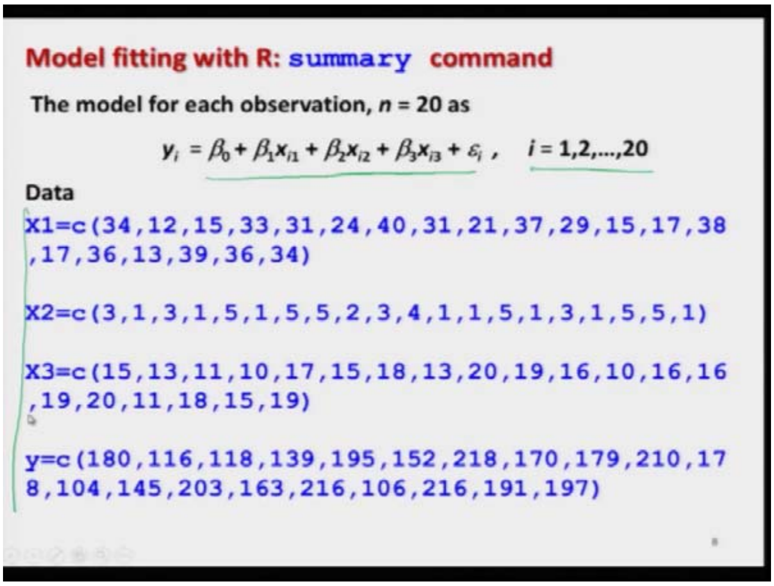


So now, for conducting the test of hypothesis, what we are going to do? We are going to use the command summary in the lm.

lm if you remember, we had used lm 2 with the linear regression model and whatever is the outcome of the lm, we will try to use the summary command on that outcome, which is obtained as an object, right. And for lm we already have discussed, that the command is lm and inside the parenthesis, you have to give the formula, you have to specify the model, and then you have to specify the data.

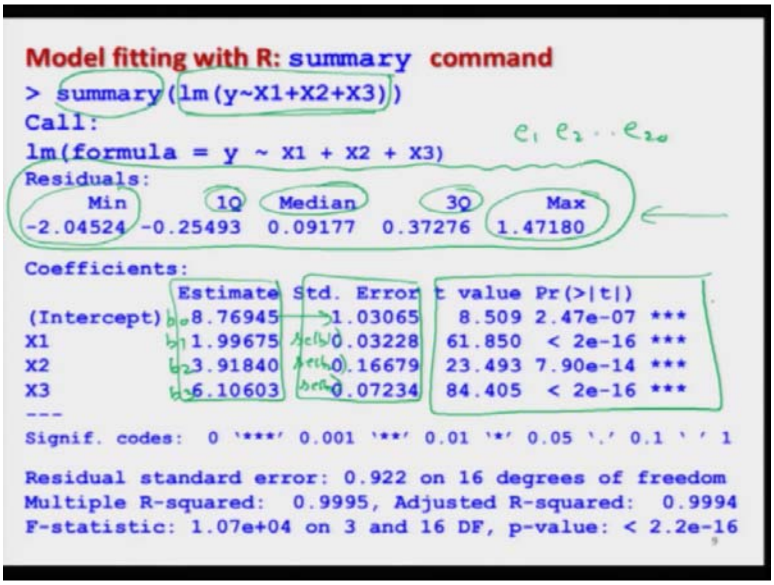
There are other commands also, but we are not going to talk about them, ok. And then whatever is the outcome here that is actually called as an object. So we try to use the summary command on this object. And then, there will be lots of outcome and we will try to see which part is indicating the test of hypothesis part.

(Refer Slide Time: 15:01)



So, I am going to consider here, the model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_i$, and we have got here 20 observation. So, I have created three data vectors for X_1, X_2, X_3 and one data vector for y, which I have stored here and you can also give it in the framework of a data frame, so that depends on you, what you want to do, right.

(Refer Slide Time: 15:31)



So, first I will try to show you this analysis on my computer and then I will try to bring you on the R console. So, now, this is your familiar slide we have used it couple of times earlier also that we have fitted here a model using the command lm and then whatever outcome is obtained from here.

I try to use the command summary over it, and then this was here is the total outcome. And, you can recall that earlier we had talked about this part is going to give you the value of b_0 , b_1 , b_2 , b_3 . So, this is b_0, this is b_1 , this is b_2 , and this is b_3 , and this column, this is trying to give you the standard error standard error of b_0 and standard error of b_1 and standard error of b_2 and standard error of b_3 .

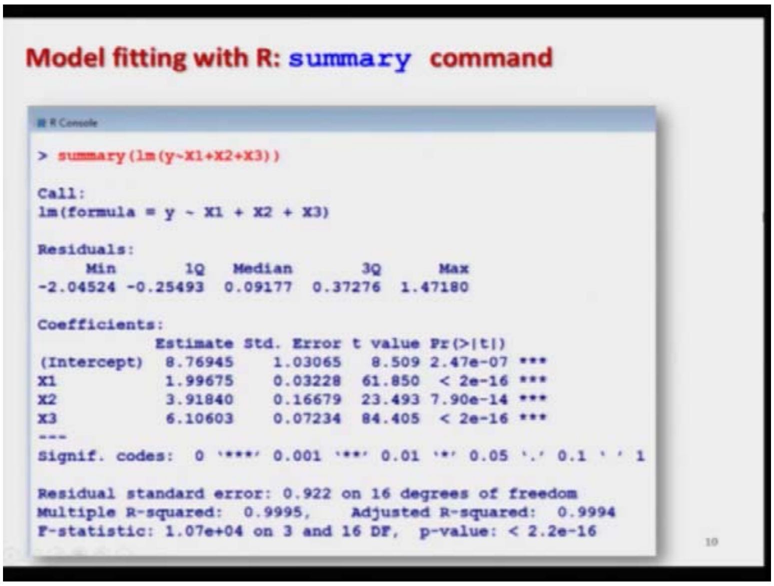
Now, we are going to considered about this aspect, right. But, before that you can also, now look here this aspect. I have not discussed about it up to now, but in the last time we had obtained the residuals, in the last lecture. So, you can see here, whatever residuals you have obtained here from say e_1 , e_2 ,…, e_{20} , this output is simply trying to give you a sort of distribution, in the sense that it is trying to give you what is the minimum value of the residual.

What is the maximum value of the residuals and what is the first quartile, second quartile which is median and third quartile of the residuals. So, whatever residuals you have obtained, so this is the simple data about it. And yeah, this will simply give you some idea that how is the distribution.

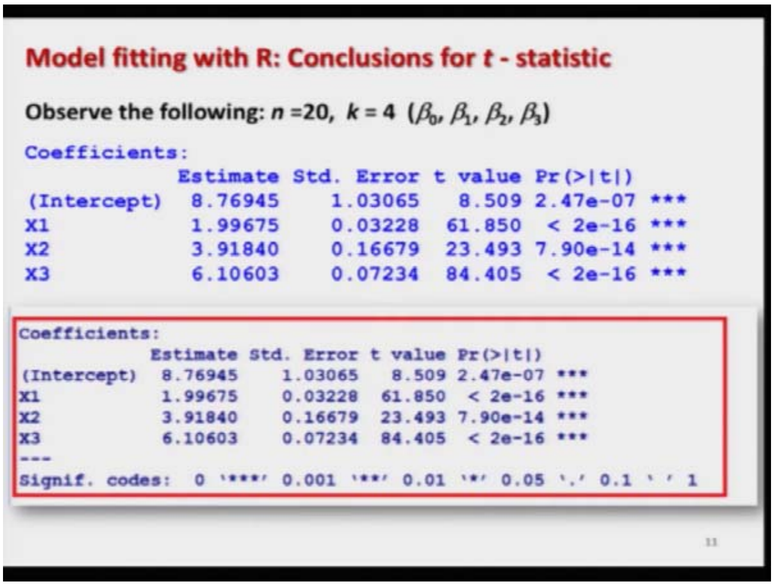
Because you are trying to assume that εs are following a normal distribution and there are various types of assumptions that we try to test. Since we cannot observe ε, so we try to take the help of residuals to visualize them, and based on that we try to take different types of decisions.

And this type of information, what is given here, this helps us in taking those types of decisions, right. So, now we are going to talk about this thing.

(Refer Slide Time: 17:50)

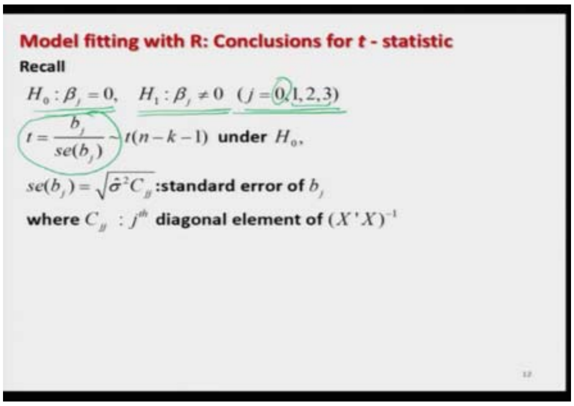


(Refer Slide Time: 17:59)



So, what I try to show you here, this is the screenshot. So, essentially now I can say once again we are going to consider on this part, ok. So now, so I have just copied here the part which I am going to consider just for the sake of clarity so that you do not get confused, right. So, you can see here this is the part, now which we are going to consider, right.

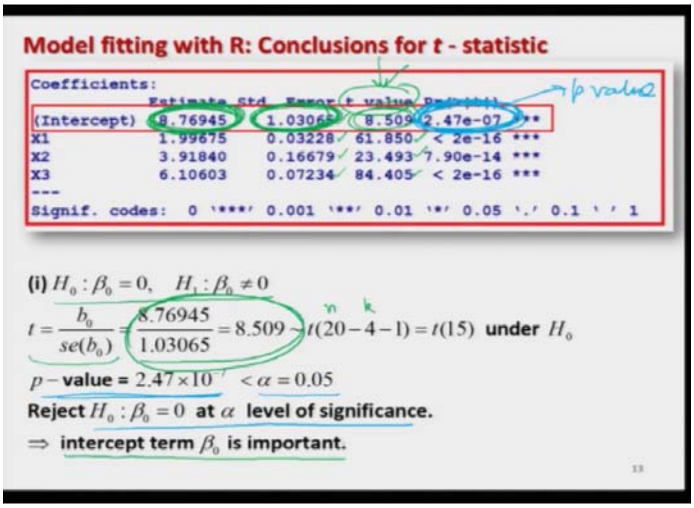
(Refer Slide Time: 18:14)



So, now just for your recollection, we are considering here the $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$, and now j is going from 0, 1, 2, 3, j=0 stands for intercept term and j=1, 2, 3 stands for regression parameters β_1 β_2 and β_3 .

And we are going to test them by t statistics. So, this is the t statistics which will be computed for j=0, 1, 2, 3. So, there should be four values of t statistics, right, which are going to be obtained. So, if you try to see here in this outcome.

(Refer Slide Time: 18:52)



So, if you try to look here t values, here you can see here there is 1, 2, 3, 4. The first one is for intercept term, then second is for β_1 , third is for β_2 and fourth is for β_3 . So, let us try to consider one by one. So, first I try to consider the test of hypothesis $H_0 : \beta_0 = 0$

against $H_1 : \beta_0 \neq 0$. So this hypothesis can be tested using the test statistics b_0 upon standard error of b_0.

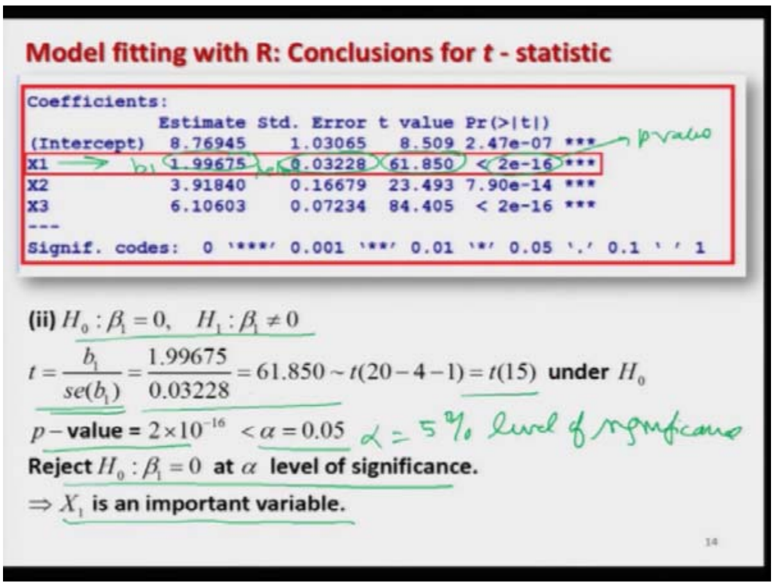
Now, if you see, what is here b_0? This is here b_0, you can see my pen. And, what is your here standard error of b_0, if you remember? This is the standard error of b_0, right. But, you do not need to compute it yourself, what I am trying to show you here that you must know that how a value has been obtained. So, if you try to see here this value here t, this has been obtained here like this, right.

So, and this is going to be n = 20, k = here 4, so the total degrees of freedom are going to be 20 - 4 - 1 which is 15. So, this t statistics has got a t distribution with 15 degrees of freedom under H naught. And, you can see here now I will use the different color pen say blue, if you try to see here this thing.

So this is here the p value, right. So, p value is written here, so you can see now here that this p value is something like 0.000000247. So this is much much smaller than the value of α. So, we can take a conclusion that the $H_0 : \beta_0 = 0$ is rejected at α level of significance.

And hence, I can conclude that yes, intercept term β_0 is important and it is contributing in my model, ok.

(Refer Slide Time: 20:55)

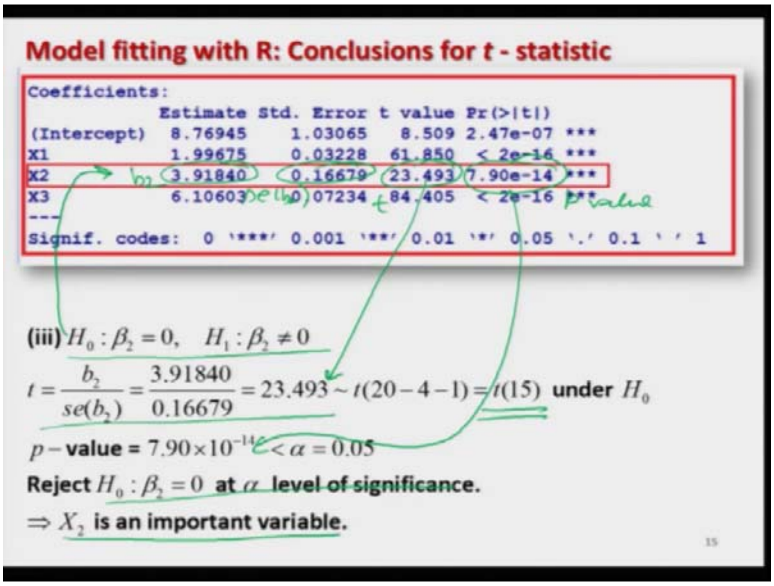


Now, the same process I can repeat for β_1 , β_2 , β_3 . So now, I try to consider here in the second line which is related to β_1 . So, this is the estimate of β_1 , which is the value of b_1 , this is the value of standard error of b_1 and this is the value of your t statistics corresponding to b_1 and this is here the p value corresponding to the null hypothesis.

So, in order to test $H_0 : \beta_1 = 0$ versus H_1 $\beta_1 \neq 0$, we use the same statistics $t = b_1 / se(b_1)$ which is obtained here like this, and this also has got a t distribution with 15 degrees of freedom and the corresponding value of $p = 2 \times 10^{-16}$ which is much much smaller than the values of $\alpha = 0.05$.

So, we are essentially considering here the 5 percent level of level of significance, right. So, α here is 5 percent. So I can say here, that reject H_0 $\beta_1 = 0$ at 5 percent level of significance. This means, that X_1 is also an important variable and X_1 is contributing in explaining the variation in y. So, this X_1 will remain in the model. That is an important variable.

(Refer Slide Time: 22:12)



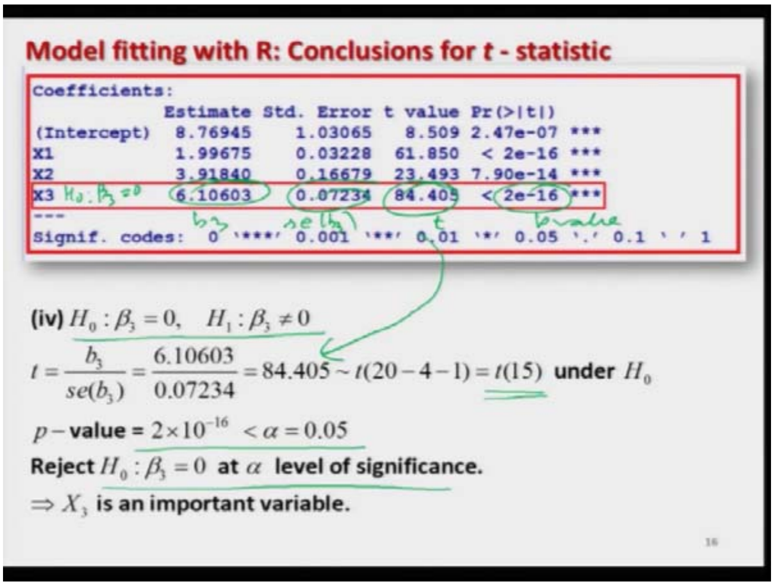
And similarly, if we try to go for $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$ and the results for this hypothesis are given in the row for X_2 .

So, this is the value of b_2 , this is the standard error of b_2 , and this is here the t value corresponding to b_2 , and this is here the p value corresponding to b_2 . So, you can see here that the t value comes out to be here 23.493 you can see here.

And, you can also obtain manually, and this will also follow a t distribution with 15 degrees of freedom under H_0 and the corresponding p value here is given by this 7.9 into 10 to the power - 14, which is very small than the value of $\alpha=0.05$.

So, this hypothesis $H_0 : \beta_2 = 0$ is also rejected at 5 percent level of significance; hence, I can say that X_2 is also an important variable and this variable has to remain in the model, ok.

(Refer Slide Time: 23:14)

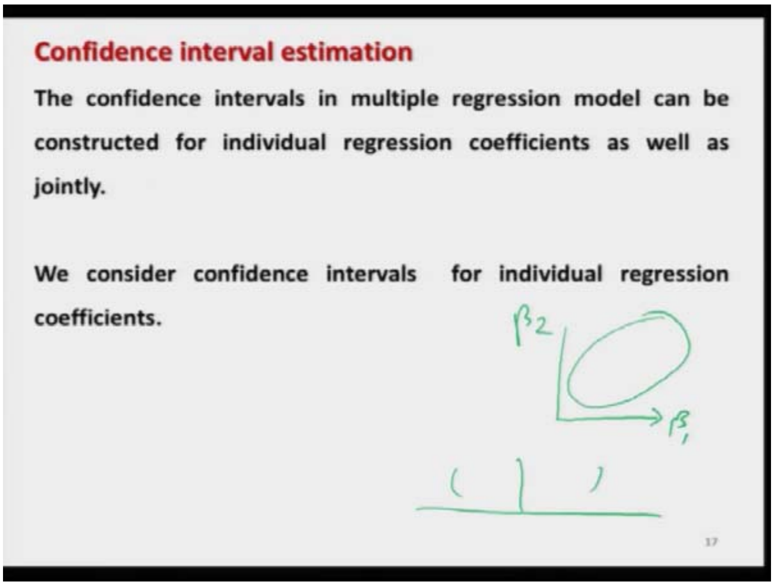


That is going to give us an important information. Finally, I try to test the hypothesis exactly on the same line about β_3 . So, H_0 is $\beta_3 = 0$ versus $H_1 : \beta_3 \neq 0$, and the information related to $H_0 : \beta_3 = 0$ can be obtained from the last column of this outcome related to X_3 .

So, X_3 means, this is about the regression coefficient associated with X_3 which is β_3 in our notation. So, again this is the value of b_3 , this is the value of standard error of b_3 , and this is the value of t statistics and this is the value of p value.

So, you can see here that this t statistics has been obtained here by 84.405 which follows a t distribution with 15 degrees of freedom, and the corresponding value of p is here 2 into 10 to the power of - 16 which is smaller than the value of α = 0.05. So, H_0 β_3 =0 is also rejected at 5 percent level of significance. So this X_3 is also an important variable which is contributing in the model.

(Refer Slide Time: 24:19)



So, now you can see here, you have identified that which are the variable. Fortunately, in this case, all the variables are important, but after that you will see that I will try to consider a topic on the variables collection and there you will see, that in the example which I have considered that all the variables are not going to be selected, all the variables cannot be considered as important.

There will always be some variable which are contributing more some, are contributing less, so we have to take a logical decision based on the statistical rule, that which of the variables are important and which are not, right, ok. So after this, test of hypothesis let me come to the confidence interval estimation.

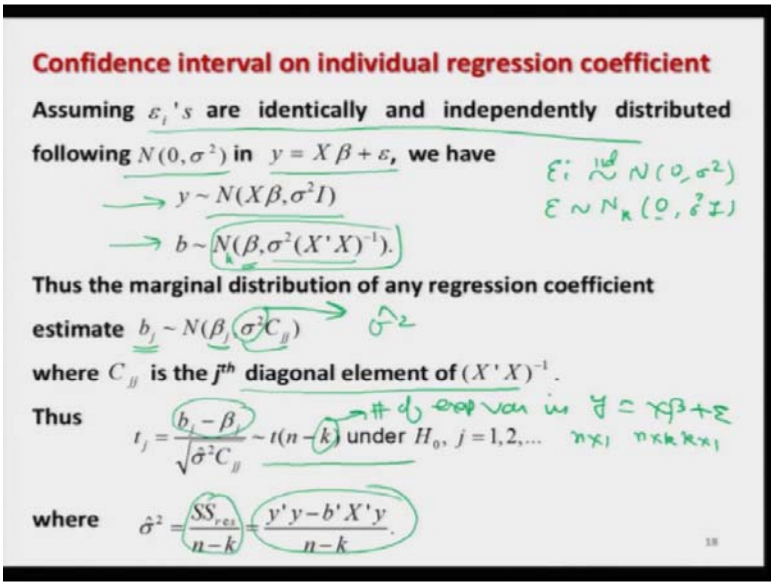
But once again I will say that I already have explained you the concept of confidence interval estimation in the case of simple linear regression model, so here I will not spend much time on the explanation of the concept. And, I also had explained that how can you construct the confidence interval. So the same methodology I am going to follow here also, right.

So now, we are going to consider here the confidence interval for the individual regression coefficient. But, just for your information, the confidence interval can also be constructed for more than one regression coefficients here, which are called the simultaneous confidence interval.

For example, if you try to take here say here, two parameters, β_1 and β_2 , then the confidence interval will be taking a look like an ellipsoid or ellipse, and if it is goes into 3 direction say, three parameter then it would be like an ellipsoid and obviously, when it is in 1 the direction this is an interval.

So, we consider now here the confidence interval for the individual regression coefficients, ok.

(Refer Slide Time: 25:57)



So, we assume here that ε i's are IID; that means, they are identically and independently distributed following a normal distribution with mean 0 and variance σ^2 in the model $y=X\beta + \varepsilon$. And, if you want to write about ε that will be a multivariate normal distribution k dimensional multivariate normal distribution with mean vector 0 and covariance matrix $\sigma^2 I$.

So, that is the same thing, whatever you like. So, based on this assumption, I can write down that we already have found that, y will follow a normal distribution with mean X β and covariance matrix σ^2 I. And hence, we also have found that b will also follow our normal distribution with mean expected value of b which is β and the covariance matrix of b is $\sigma^2 (X'X)^{-1}$, right.

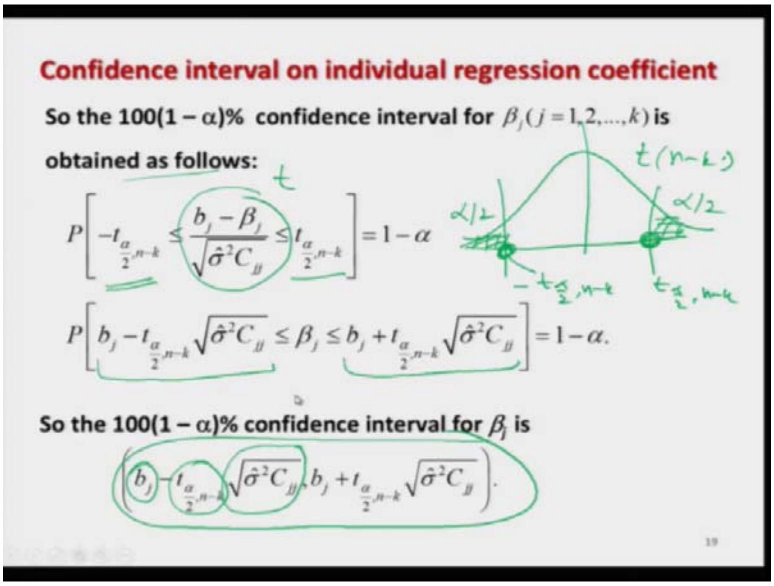
And, if you try to find out the marginal distribution of any regression coefficient from this one, because this is here a multivariate normal distribution of k^{th} dimension. So, if you try to find out the marginal distribution of any particular estimate say b_j, so that is going to be a univariate normal distribution with mean β_j, and say this variance $\sigma^2 C_{jj}$, where C_{jj} is the j th diagonal element of the matrix $(X'X)^{-1}$.

So, now based on that, we know that how to construct the t statistics. So, that is going to be $(b_j - \beta_j)/se(b_j)$. So, standard error of b_j can be obtained by replacing this σ^2 in the covariance matrix of b_j by $\hat{\sigma}^2$

So, I can write down the $t_j = \frac{b_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t(n-k)$, and this is in general going to follow a t distribution with n - k degrees of freedom, right. So, if you remember that here, k is the number of explanatory variables in the model $y = X\beta + \varepsilon$.

Where we have assumed that y is a n x 1 vector and X is a n x k matrix and β is a k x 1 vector, right. So, this is what you have to always keep in mind. And, this $\hat{\sigma}^2$ is going to be obtained from the expression, sum of square due to residual divided by n - k, and the alternative expression is given by like this. So that we already have discussed.

(Refer Slide Time: 28:33)



So, now you can see that whatever concept, we have learnt up to now, we are trying to comprehend them and we try to use them. So now, if you go back to your lectures in the case of simple linear regression modeling and try to recall how we had constructed the confidence interval, then the same approach I am going to use here also.

So, the 100 1 - α percent confidence interval for β_j can be obtained as follows, right. Try to write down here the t statistics, and t statistics is going to follow a t distribution and on the left hand side and, right hand side of the t distribution with n - k degrees of freedom.

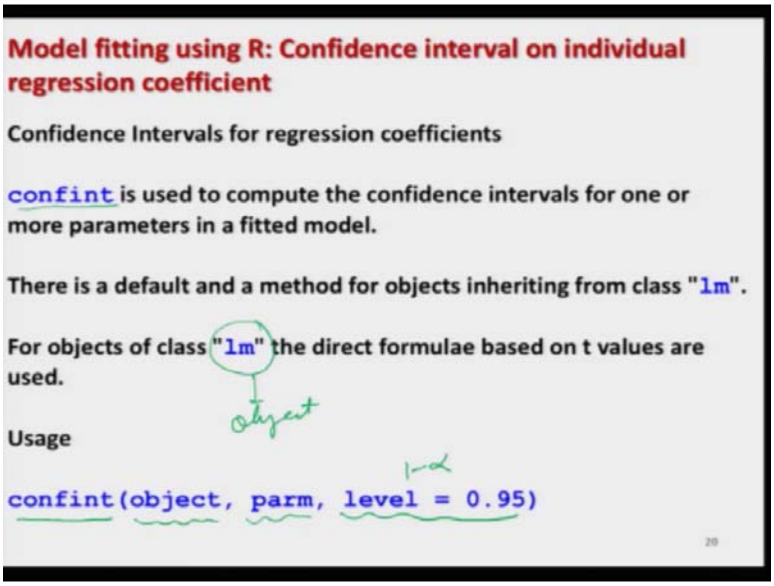
I try to denote this area which is shaded here as α/2 on the left hand side and shaded area on the, right hand side as α/2. And, corresponding to which here, there will be two critical values. So this value is going to be - $t_{\alpha/2}$ and the value on the, right hand side will be $t_{\alpha/2\ n-k}$ and somewhere it will be the mean, ok.

So, now you assume that this statistic t lies between $-t_{\alpha/2\ n-k}$ and + $t_{\alpha/2\ n-k}$, and you simply try to solve this inequality and you can obtain that, this is here the lower limit of the confidence interval and this is here the upper limit of the confidence interval. So you can very easily solve it and hence, the 100 (1 - α)% confidence interval for β_j can be obtained here like this, ok.

So, this is how you can obtain the confidence interval. Which is very simple straight forward. Now, this b_j is known to you, $t_{\alpha/2\ n-k}$ can be obtained from the table, that can be also be obtained as a percentile directly from the R software. These values they are available from the output of the software.

So, this lower limit and upper limit can be computed manually also very easily, but definitely we are going to use here the software.

(Refer Slide Time: 30:27)

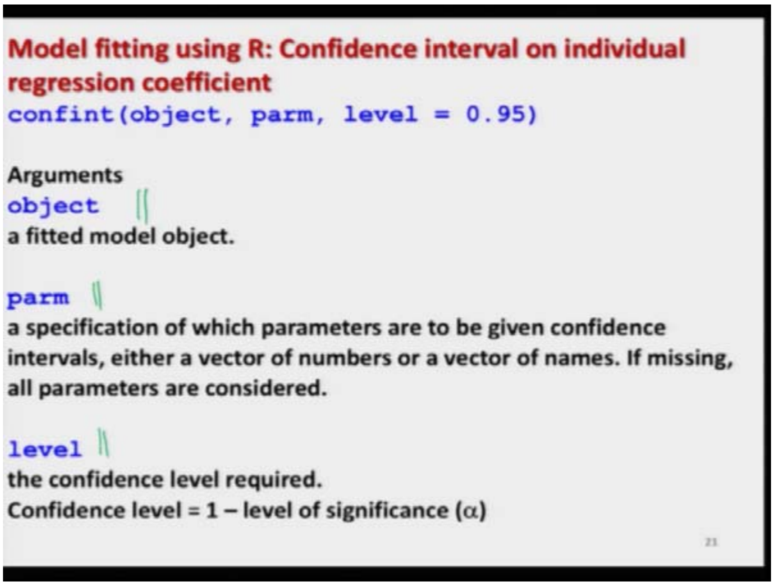


So, now in the R software how would you use it that is the next objective. So, we try to use command conf int. If you try to see, this is the same command that we had used earlier also. So, in this case what we try to do? First, we have to use the command lm and whatever is the outcome of this lm that is stored as an object and from that object, I have to extract the confidence interval of this individual regression coefficient.

So, the command to find out the confidence interval goes like this, try to write down the command conf int, then try to write down the object, then try to write down the parameter, for which you want to have the confidence interval, and if you do not give any name of the parameter.

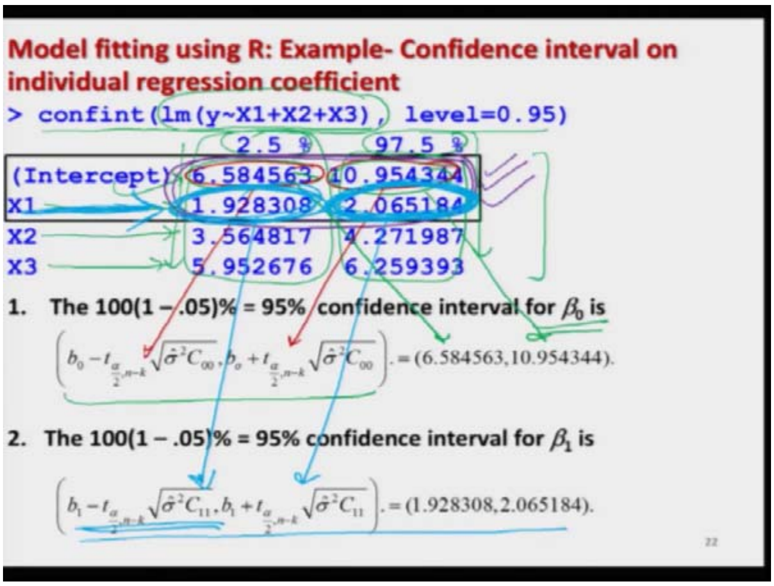
Then all the confidence interval corresponding to all the parameters will be given in the output. And then you have defined here the level. So, remember level is defined here as say 1 - α. So if you are trying to take 5 percent level of significance, then the level will be 0.95 that is 95 percent.

(Refer Slide Time: 31:30)



So, this is how we are going to do. So, this is the explanation of this command. So this is about object, this is about parameter, and this is about level. So, this is just for your information so that you can when you try to read this then you can recall all the things, ok.

(Refer Slide Time: 31:46)



21

So, now let me try to use this command on the R console, and first I try to show you here the outcome and then I will try to show you on the R console, ok. So, now, first I try to consider here the confidence interval you can see here, I am trying to use here conf int and then inside the parenthesis.

This is my object, this is the command to obtain the fitted linear regression model, and then I am trying to say here level is equal to 0.95; that means I want 95 percent confidence interval. And I am not using here the parameter or the option p a r m, because I want all the confidence interval related to all the parameters. So, you can see here this will be the outcome.

And, you can see here this is here 2.5 percent and this is here 97.5 percent and then there are here 4 values and 4 values here. So, this 2.5 percent is indicating actually the lower limit of the confidence interval related to intercept term, related to X_1 which is β_1 for β_2 and for β_3. And similarly, this part here is the upper limit, which is the 97.5 percentile of the given data.

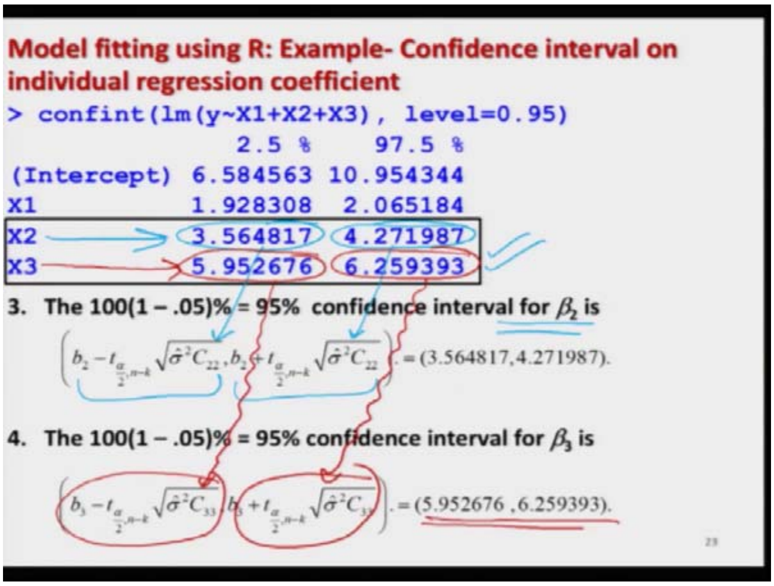
So, now the thing is this, we try to first understand these four outcomes. So, first we try to consider this part which I have enclosed inside a rectangle, black rectangle. So this is about outcome you can see here I am highlighting it, ok. So we are going to concentrate on this part. So, first we consider the construction of the confidence interval, 95 percent confidence interval for β_0.

So, you can recall that we have constructed the confidence interval for β_0 like this, where the lower and upper limits are given by these two expressions. So, you can compute these values manually also, but in the software you can see, this is here the value. This is here the lower value and this is here the upper value, or I try to use here a different color pen say red color pen.

So this value here is for the lower limit and the value of the upper limit, right. And similarly, if you try to find out the 95 percent confidence interval for the β_1 then the interval is given by here like this and this value has been obtained here. Try to use this movement of my pen in blue color and this is the lower limit of the confidence interval which is obtained by this expression.

And this second value we try to look at moment of my pen, this is obtained here by this expression which is the upper limit of the confidence interval for β_1 . So, this is how you can see, that these are the values and this is how they are computed in the software.

(Refer Slide Time: 34:40)



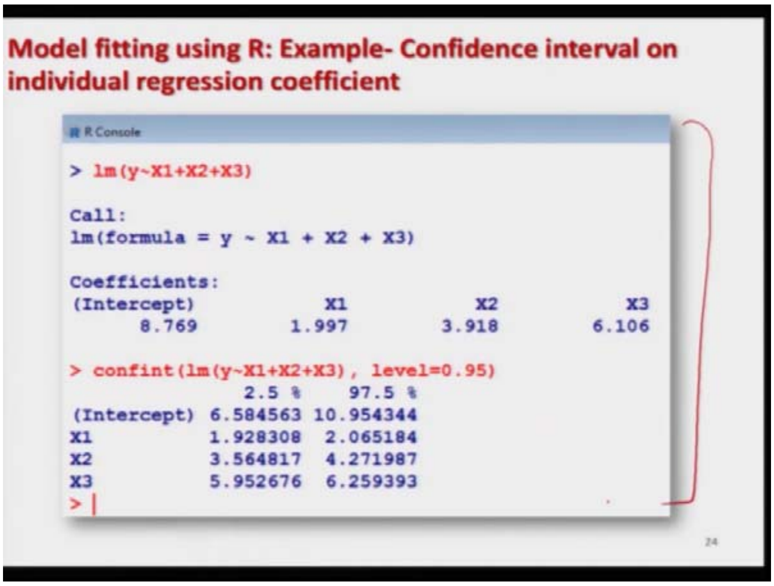
And similarly, if you try to concentrate on the confidence interval for β_2 and β_3 , try to concentrate on this part which is here inside the black, say, box, right. So, if you try to see here the 95 percent combination for β_2 , this is here like this try to look into the column or the row of the X_2 .

So, this is here the value of the lower limit which is obtained by using this expression. This is the formula for the lower limit and this is the second value here is the value of the upper limit corresponding to β_2 , and this is the value of the second that is the upper limit of the confidence interval.

And similarly, for this β_3 , this value here X_3 is trying to give you β_3 . So, this is essentially the lower limit of the confidence interval and this is here the upper limit of the confidence interval.

And this limits have been obtained the lower limit has been obtained by using this expression and the upper limit has been obtained by using this expression which is the upper limit of the confidence interval. And you can see here, that these are the 95 percent confidence interval in which the β_3 is expected to lie, right.

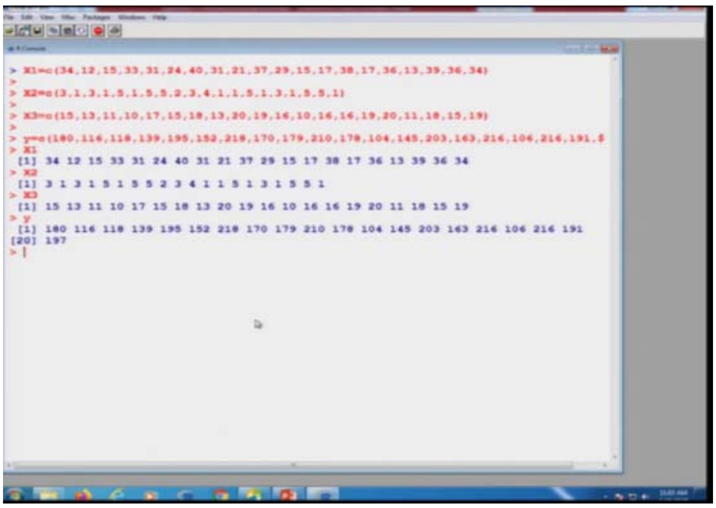
(Refer Slide Time: 35:54)



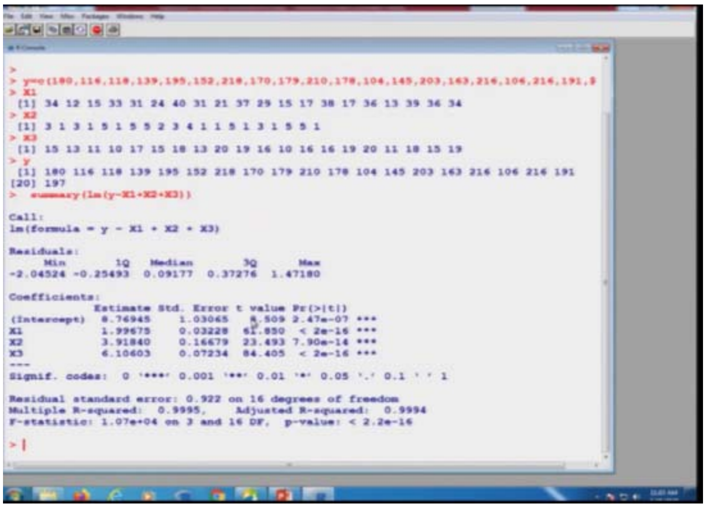
So, now this is the screenshot of the same thing what you have seen up to now. And, I will try to now show you all these things on the R console also. So, first let us try to consider the test of hypothesis.

So, you can see here we have this data. So, I already have entered this data in the R console, I will show you and then we have to use this command here on the R console.

(Refer Slide Time: 36:30)



(Refer Slide Time: 36:42)



So, you can see here that I already have entered this data. You can see here, this is X1, this is X2, this is here X3 and this is this is here y. Now, if you remember, once you try to fit here model and then try to find out its summary command, it will look like this, right.

So, you can now see at this part, this part here, right. So, this part whatever I have given here, this is the this thing which I will try to show you here, right. So, this is exactly the same thing. So, what you have to see here? That you try to look at this value, ok. I will just highlight it and you try to see the movement of my cursor.

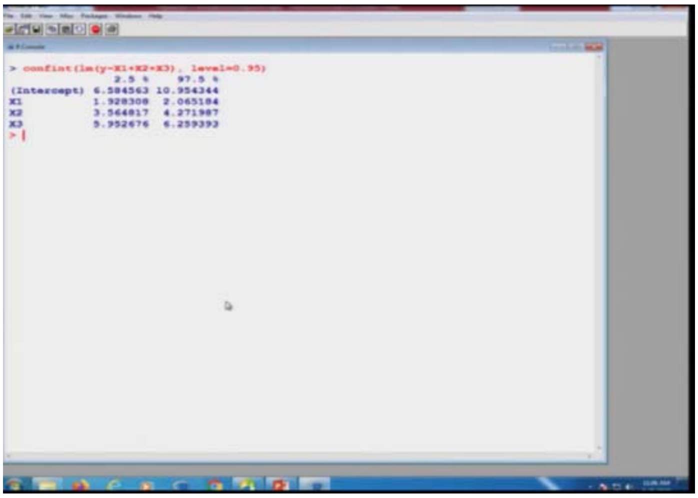
So, this value here is the t value for intercept term, and this is here which I am highlighting now, this is the p value corresponding to $H_0 : \beta_0 = 0$. And similarly, if you try to observe this highlighted value this is the t value corresponding to $H_0 : \beta_1 = 0$ and this is here is the p value which is the corresponding to $H_0 : \beta_1 = 0$.

Similarly, this value highlighted here is the value of t statistics corresponding to H_0 $\beta_2 = 0$ and this is the corresponding p value. And similarly, this last value which I am highlighting, this is the value of the t statistics corresponding to $H_0 : \beta_3 = 0$ and this is the here is the p value which is corresponding to this hypothesis $H_0 : \beta_3 = 0$.

And yeah means; obviously, this significance code I they are trying to give you here, the value of α and then this these 3 stars are going to indicate at what level of α they are being considered. And we had already I mean discussed these things when we had done the simple linear regression model. So I will skip that part.

But you can see here, that finding out the conclusion about the test of hypothesis is not difficult at all, right. So, after this I will try to obtain here the confidence interval. So you can see here that we already have obtained this summary command, and we already have found the lm command.

(Refer Slide Time: 38:54)



So, I will try to use here this command and you can see here once I try to use it, I will clear the screen so that you can see it very clearly. You can see here, this is my confidence interval.

So, this is the lower limit of the confidence interval for β_0, and this is the upper limit of the confidence interval for β 0. Try to just watch where I am trying to highlight on the screen. And, this is the lower limit of the confidence interval for β_1, and this is the upper limit of the confidence interval for β_1.

And similarly, this is here is the lower limit for the confidence interval for β_2 and this is here is the upper limit of the confidence interval for β_2. And similarly, this is here the confidence interval for β_3, lower and upper limit, right.

So, you can see here, it is not very difficult to find out the confidence interval in case of this multiple linear regression model in the R software, yes, ok now. So, the time has come to stop in this lecture. And, I have given you the details about the test of hypothesis and confidence interval estimation for the individual coefficients, right.

I am not considering here the simultaneous confidence interval, but that is not difficult. Once, you have understood these things, now you can actually try yourself. And this is what I want that you should stand on your own feet. Sometime I get many emails or calls people try to ask me.

Well, I am making this mistake can you please help me, but I always say try to help yourself, because finally you have to stand on your own feet, that is the best thing in life. You should be complete and sufficient in yourself. You should not depend on anybody else, and if you try hard there is no reason that why you cannot solve the problem.

So, that is my personal belief, and I believe that you also have the same philosophy in your life. So, now I will request you try to take an example, whatever example you have considered earlier and try to conduct the test of hypothesis and confidence interval estimation under those example and try to see what do you get.

Try to learn how to interpret it. And the more you practice, the more you will be confident, more you will learn. So, you practice and I will see you in the next lecture with another topic on Multiple Linear Regression Model, till then good bye.