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 - 43
Simple Linear Regression Analysis
Maximum Likelihood and Confidence Interval Estimation

Hello friends, welcome to the course Essentials of Data Science with R Software 2 where we are trying to understand the topics of Sampling Theory and Linear Regression Analysis and in this module on the Linear Regression Analysis, we are going to continue with our chapter Simple Linear Regression Analysis.

So, in this lecture, we are going to consider another method of estimation for the regression parameters as well as the variance. So, you may recall that up to now, we have considered the principle of least squares and we have used the direct regression method estimation to obtain the value of the regression coefficients as well as the variance, variance of the random error component.

In statistics, there are different types of estimation methods which are based on different types of concept, different types of philosophies and among them; one very popular method is maximum likelihood estimation. So, today, we are going to discuss about the Maximum Likelihood Estimation and then, we will talk about the test of hypothesis.

So, first what is this maximum likelihood estimation? You see likelihood means what? The literal meaning of likelihood is what? It is related to the probability that what are the chances. So, you are trying to say maximum likelihood mean what are the maximum chances, chances of what? So, we are trying to draw a sample from some population. So, now, we need to assume a form of the population. So, we will assume that suppose your sample is coming from a normal population with some finite mean and finite variance.

And now, we are assuming that the sample is truly representative of the population. So, that means, whatever you expect from the population that should be present in the sample also. So, that means, now you are trying to observe the sample and you are trying to estimate the most probable value of the parameter that may happen right.

So, in order to obtain the maximum likelihood estimation, what we try to do? We try to maximize the probability and we try to obtain the values of the parameters which are most probable to occur, and this is the basic idea about the maximum likelihood estimation.

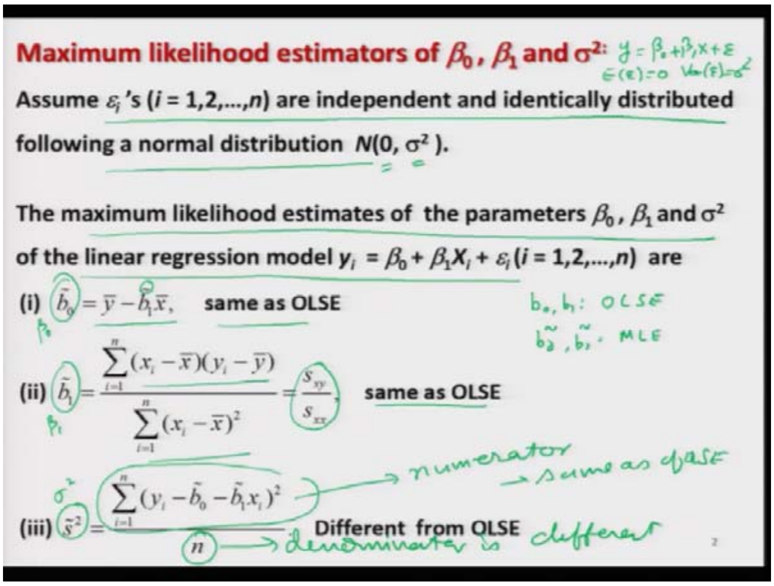
So, what we have to do? We simply have to write down the probability density function of all the observations and then, we have to maximize it because the probability distribution functions are going to give us the probability of occurrence, probability of occurrence of certain events.

So, now, we are going to maximize this probability and we will try to find out the values of the parameter for which this probability will be maximum. So, now, if you try to see once you talk about maximization so, in finite sample, I can assume a nice probability distribution in which means we can do some algebra and we can find some exact expression of the maximum likelihood estimators.

But in data sciences, what you can do? That you can use different types of algorithm also, if you are dealing with some complicated distribution which is not necessarily normal so, you can consider those thing, you can employ different types of optimization technique and a basic objective is this you simply have to maximize the likelihood or maximize the probability right.

So, first we try to talk about the maximum likelihood estimation and then, I will give you the brief introduction to the test of hypothesis. Although, I am assuming that you have a sufficient background in statistics. So, all of you are familiar with the concepts of maximum likelihood estimation and test of hypothesis. So, let us begin our lecture from this slide right ok.

(Refer Slide Time: 04:31)



So, now, we are going to consider the maximum likelihood estimation of the parameter β_0, β_1 and σ^2. You can recall that we had taken a model $y = \beta_0 + \beta_1 x + \varepsilon$ and ε had mean 0 and variance of ε was σ^2. So, these are the three parameters in which we are interested.

So, now, we are making one assumption that we are trying to associate a probability density function with epsilons because unless and until you associate a probability density function, you cannot write down the likelihood function. So, we assume that ε i's are identically and independently distributed following a normal distribution with mean 0 and constant variance σ^2, right.

So, first I try to give you the final outcome and then, I will try to show you how to get them. So, when we try to find out the maximum likelihood estimators of β_0, β_1 is σ^2 in the linear regression model, this $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, i goes from 1 to n, then the maximum likelihood estimator of the β_0 intercept term turns out to be $\bar{y} - \tilde{b}_1 \bar{x}$.

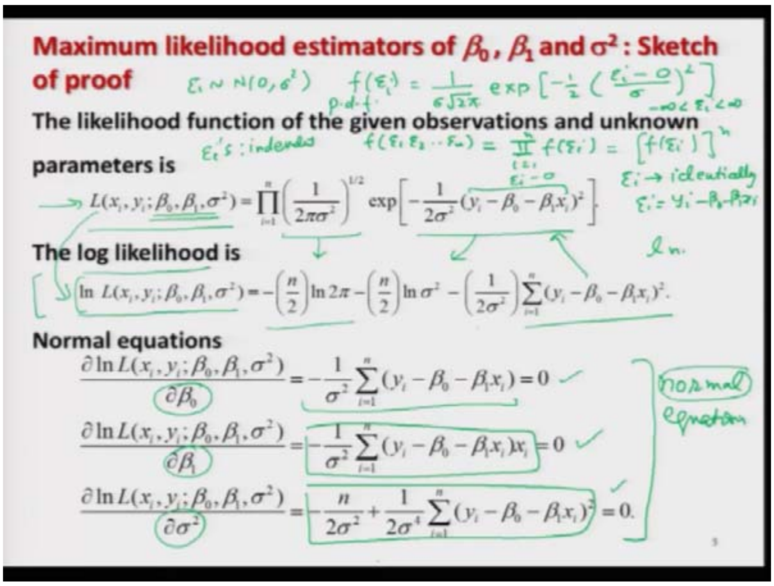
So, you can see here, I am using here symbol here $\tilde{b}$. So, that is really going to indicate that this estimator is different from b_1. So, we had earlier denoted b_0 and b_1 for ordinary least square estimator and we will denote $\tilde{b}_0$ and $\tilde{b}_1$ for maximum likelihood estimator.

So, the estimator of β_1 slope parameter turns out to be $\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{s_{xy}}{s_{xx}}$. So, you can recall that these two estimators of β_0 and β_1 which are the maximum likelihood estimators of β_0 and β_1, they are the same as ordinary least square estimator right ok.

Now, if you try to estimate the variance σ square, then the maximum likelihood estimator of σ^2 comes out to be summation y i – b 0 tilde – $\tilde{b}_1$ x i whole square divided by n and this we are denoting by here $\tilde{s}^2$. Now, you can see this is different from the ordinary least square estimator of σ^2 what is the difference?

You can see here this $\frac{\sum_{i=1}^{n}(y_i-\tilde{b}_0-\tilde{b}_1x_i)^2}{n}$ numerator part is; numerator is same as of OLSE, but if you come to denominator, denominator is different what is the difference? In case of ordinary least square estimator, this denominator was n – 2, but here it is n. So, the maximum likelihood estimators and ordinary least square estimator of β_0 and β_1 are the same whereas, the ordinary least square estimator of σ^2 and the maximum likelihood estimator of σ^2, they are different.

(Refer Slide Time: 08:22)



So, now, we try to see how these estimators have been obtained. So, we try to have a sketch of the proof of maximum likelihood estimators. So, one thing you have to remember that when you are writing that ε_i is following a normal distribution with mean 0 and variance σ^2, then the probability density function of ε i will be given by $\left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left[-\frac{1}{2\sigma^2}(y_i-\beta_0-\beta_1x_i)^2\right]$, right where your ε lies between $-\infty$ and $+\infty$.

So, this is actually the probability density function of N(0, σ^2). Now, you are assuming that ε_i's are independent. So, when they are independent, then the joint density function of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ this can be written as the product of the individual probabilities probability density function i goes from 1 to n and once you are assuming that ε i's are, they are identically distributed, identically means all $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ they have got the same distribution with mean 0 and variance σ^2 right.

So, this can be written here as say $\left(f(\varepsilon_i)\right)^n$, right. So, this is the basic concept which I am going to use here right. And you can see here that this ε_i is in is your $y_i - \beta_0 - \beta_1 x_i$ so, that is what I have written here. I have written the joint density function of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ in terms of x_i, y_i, β_0, β_1 and σ^2.

So, you can see here this is the same thing, this is the $\prod_{i=1}^{n}\left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left[-\frac{1}{2\sigma^2}(\varepsilon_i-0)^2\right]$, right. So, what we try to do here that we have to now maximize this likelihood function and we have to obtain the values of β_0, β_1 and σ^2.

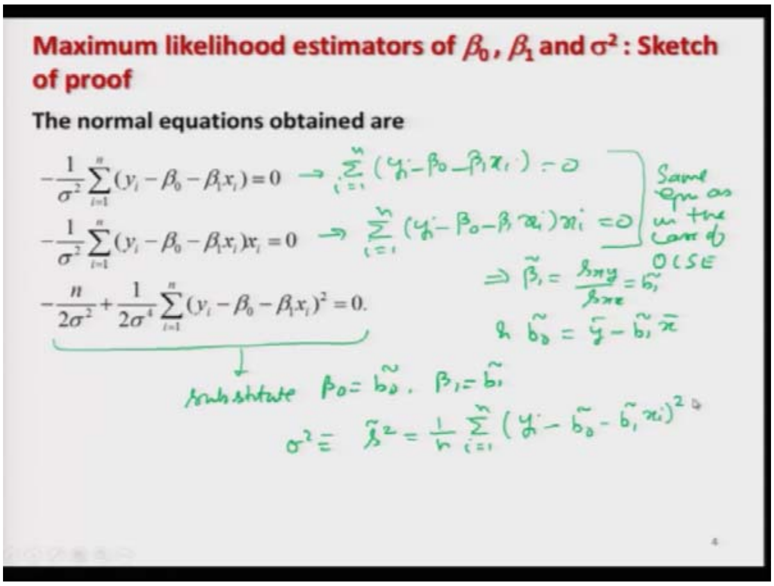
So, we know that this likelihood and the log of likelihood both are monotonic functions. So, instead of maximizing the likelihood function, I can maximize the log likelihood also. So, that is going to give us the same outcome. So, now I consider here the log of this likelihood and I try to write down here this quantity will become – n by 2 log of 2 pi that means, I am taking here natural log, natural log is denoted by ln ok and this quantity here becomes here see here $-\left(\frac{n}{2}\right)\ln 2\pi-\left(\frac{n}{2}\right)\ln\sigma^2-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2$ and then, this quantity from here ok.

So, now this is the log likelihood which I have to maximize. So, I try to use here the principle of maxima and minima and I try to partially differentiate this log likelihood with respect to β_0, β_1 and σ^2 and once I try to differentiate the log likelihood with respect to β_0, I obtain this result that you can verify, that is very straight forward.

And similarly, when I try to partially differentiate log of likelihood with respect to β_1, then I get this value and similarly, when I try to differentiate the log likelihood or partially differentiate the log likelihood with respect to σ square, then I get here this value. Now, I try to equate them equal to 0 so, that is what I am doing here, all the three equations have been equated to 0 and these three equations are called as normal equations.

Remember one thing, do not get confused that this word normal is coming from normal distribution. The normal in normal distribution and the normal in normal equation, they are different thing right. Even if you try to take here any other distribution say binomial or Poisson also, even then you will have normal equations.

(Refer Slide Time: 13:25)



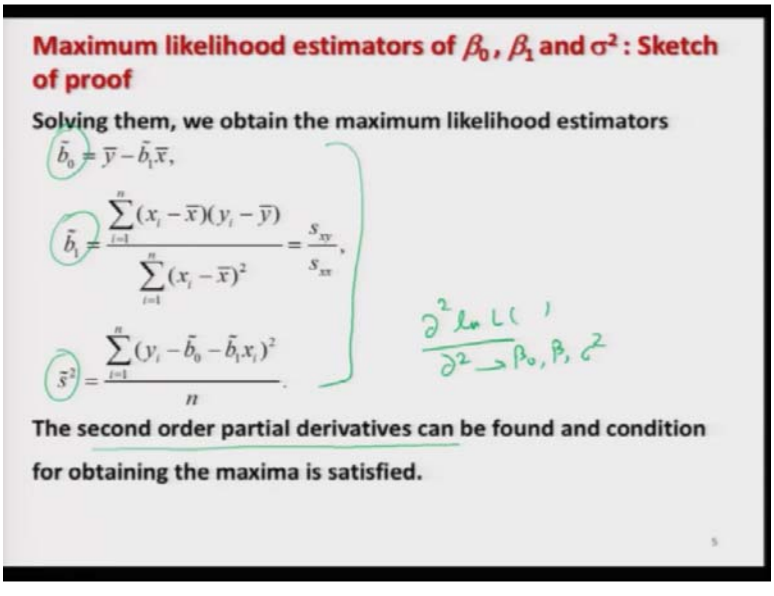
These are the three normal equation that we have obtained and if you try to see what is this thing, this can be written as $\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)=0$.

And now, can you really recall have you ever solved these equations earlier? They are the same equations as in the case of least square estimation. When you obtain the ordinary least square estimate, you solve the same equation. So, now, I do not need to solve it again right.

So, if you try to solve them, they will give you here $\tilde{b}_1$ to be s_{xy}/s_{xx} and which I have denoted by see here $\tilde{b}_1$ which is the maximum likelihood estimator of β_1 and the maximum likelihood estimator of β_0, this is $\tilde{b}_0$, this is . $\bar{y} - \tilde{b}_1 \bar{x}$.

Now, if you try to solve this equation and try to substitute β_0 equal to $\tilde{b}_0$ and β_1 equal to $\tilde{b}_1$ in this equation and try to solve here, you will get here the value of σ^2 that will be your estimate which is denoted as $\tilde{s}^2 = \frac{\sum_{i=1}^{n}(y_i - \tilde{b}_0 - \tilde{b}_1 x_i)^2}{n}$. So, you can obtain these things.

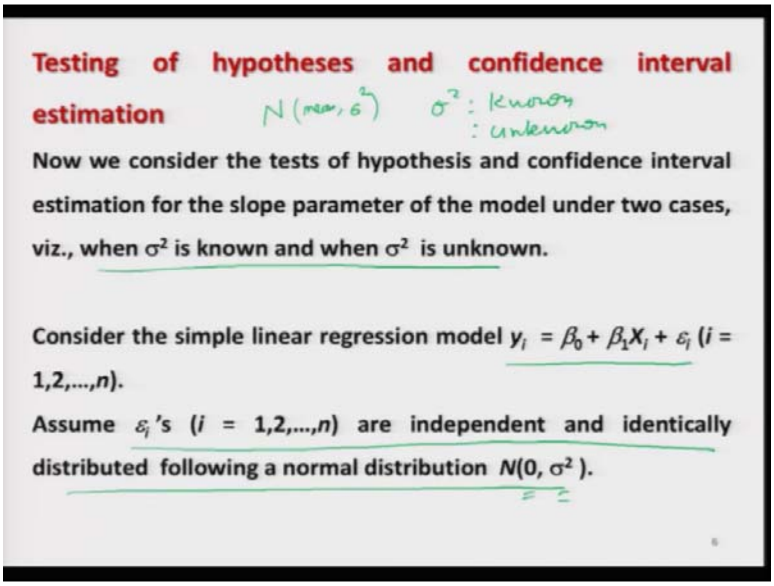
(Refer Slide Time: 15:36)



And yeah, so, these are the final expression that you will get. So, these are the values which are expected to maximize the likelihood function, but how to cross check it? So, for that I will not do it here, but I will try to do it in more detail when we will try to consider the multiple linear regression model, but here, you can find out the second order partial derivatives of the log likelihood something like you have find out the log of L, try to find out the second order derivative with respect to β_0, β_1, σ^2 from where, from here right.

From these three equation, you try to differentiate them, once again you will get the second order partial derivatives and then, try to create the bordered Hessian matrix and then, you can show that these three values of β_0, β_1 and σ^2 that you have obtained, they are maximizing the likelihood function. Hence, these three values can be call as the maximum likelihood estimators of β_0, β_1 as well as σ^2 right ok. So, this was all about the maximum likelihood estimation.

(Refer Slide Time: 16:53)



Now, I try to come on different aspect which is test of hypothesis right. So, before I do, let me try to give you some idea. So, the first question is why do we conduct the test of hypothesis?

And sometime, people get confused between the estimation and test of hypothesis. Now, you see you have employed two different estimation methods to estimate the same parameter β_0, β_1 and σ^2. Fortunately, and just by chance, the estimators for β_0 and β_1, they come out to be the same.

The OLSE as well as MLE, they are the same for β_0 and β_1, but that is only by chance means it is not a rule that these estimators are going to be the same, but you have seen that the variance estimator for OLSE and MLE, they are different. So, now whatever estimation technique you use, they are based on the random sample. You are simply trying to find out some mean, some sum of squares and then, you are trying to compute the value of parameter.

Now, suppose the sample changes, you have seen through the theory of sampling theory that as soon as you draw a different sample, you will get different values of the statistic. So, now the question is different samples will generate different types of values of these parameters. So, the question is which one is right, which one is wrong and various types of questions crops up.

So, we assume that as long as we are getting the random samples and the difference in the values of these parameters that we are obtaining, that is just due to random variation we can accept it because we have no other option, the random variation is beyond our control. But if there is some assignable cause, that should be taken care right.

So, what we try to do here that we try to find out the values from the sample and then, we try to compare them with some given values and based on that, we try to make a comment whether the estimated value can be considered to be the same as the value which we have assumed or it is known to us that is the main objective of test of hypothesis and what I am doing is this a very layman language right, I am not explaining you here the basic concepts of test of hypothesis right that I assume that you know it.

So, one thing which I would like to clear that many times people get confused that the test of hypothesis can also give us the values of the parameter, the answer is no. The values of the parameter can be obtained only through the estimation methods, the test of hypothesis can only check whether the estimated values are equal to or less than or greater than the given values. So, the test of hypothesis is conducted against some known values.

But if the hypothesis is not accepted, the then the methodology of test of hypothesis will not help us in identifying or knowing that well, what is the correct value the question goes. Suppose you estimate something and I say ok, the suppose you estimate the sample mean to be 20 and I assume that the population mean is suppose 30 well, and if the difference is due to the random causes and suppose the hypothesis is not accepted so, that means, I will be concluding that the sample mean and the population mean are different.

Now, you can ask me next question. Ok, you are saying that ok, these two values are different that I am accepting, but please tell me what is the correct value then, then I will say sorry, the test of hypothesis cannot answer this question that you have to look into your estimation method, the properties of estimation techniques and then, you have to decide that what can be a better value which will be close to the true value ok. So, that is the basic idea.

Now, we are estimating here three quantities β_0, β_1 and σ^2. So, we would like to conduct the test of hypothesis for all these three parameters. In this chapter, I am trying to show you that how you can conduct the test of hypothesis for this single parameters and in the case of multiple linear regression model, I will try to extend this concept of test of hypothesis in casting the equality or something like this for more than one parameters also right.

And after this, I will try to consider the confidence interval. What is confidence interval? For example, you have utilized the principle of least square and method of maximum likelihood to estimate the values of β_0, β_1 and σ^2. Well, these estimation techniques are giving you one particular value, the value of these parameters at a point so, these are actually point estimate.

Point estimate means you are trying to say that ok, the value of this parameter is this, but now, the other alternative is this, I can find a probable interval in which the population parameter value will lie, right. The most simple example goes like this for example, if there is a medicine and we conduct a experiment; we conduct an experiment to know that what is the approximate number of hours by which the medicine can control the fever.

You take a sample and suppose you find out the sample mean and then, try to compute the statistics and suppose you come to know that the value of the parameter is suppose 6 hours so, that mean that medicine is expected to control the fever for 6 hours. But do you think that this 6 hours is going to be the same for all the persons?

Some person, in some person the medicine may be effective for say 5 hours 30 minutes and some persons the medicine will be effective for 6 hour 30 minutes, in some persons the time may be smaller than 6 or more than 6 also right. So, for that if I say ok, this medicine can control the body temperature or fever say from 5 to 7 hours, this also makes sense.

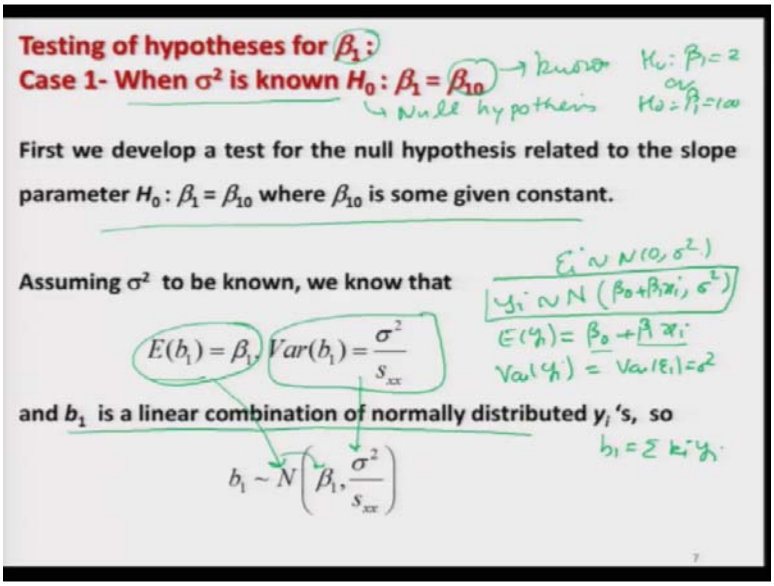
What is this? This is confidence interval estimates that means, we are trying to find out the value of the same parameter not at a point, but in the form of an interval. So, for the interval, there will be a lower limit, there will be an upper limit. So, our objective here is this, we already have now learnt that how to estimate the parameters at a point now, we will consider how to estimate them in an interval, right.

So, first we try to consider the test of hypothesis and then, we come to the confidence interval estimation and you will see at a later stage that confidence interval estimation and test of hypothesis both are actually inter related through something. So, let us now come back to our slides ok.

Now, we try to assume that we will be working under the normal population. This normal population will have some mean and will have some variance σ^2. So, we have here two possibilities that this σ^2 is known, or σ^2 is unknown. So, we will consider here two possible cases and we will try to construct the test of hypothesis and confidence interval estimation under two possible cases, when σ^2 is known and when σ^2 is unknown.

So, our model is going to be the same $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and we have got small a number of pairs of observation and we assume that this ε i's are independent and identically distributed which are following a $N(0, \sigma^2)$, ok.

(Refer Slide Time: 26:14)



So, now let me first consider the construction of test of hypothesis for the slope parameter β_1 and the first case, I try to take when σ^2 is known and our H_0 which is null hypothesis is $\beta_1 = \beta_{10}$. So, β_{10} is some known value right, I can write down here say $H_0 : \beta_1 = 2$ or $H_0 : \beta_1 = 100$ whatever you want.

So, you want to test that the values which you have obtained from the sample for this parameter β_1, are they equal to the value β_{10} in the population or not right. So, this is our now set up that β_{10} is some given constant. Now, if you remember in the case of ordinary least square estimation, we had already proved that the estimators b_1 that is the least square estimator for the parameter β_1 is an unbiased estimator and its variance was obtained like this right.

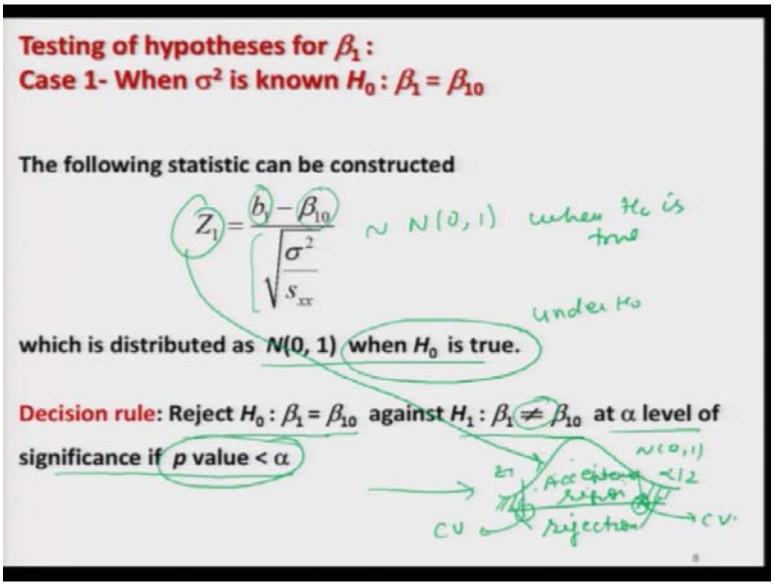
And we also had shown that b_1 was a linear combination of y i's and now, we are since we are assuming here that ε i's are following normal 0 σ^2 so, will be y_i will be also be a normally distributed random variable why? Because the linear combination of normally distributed random variable under certain condition is also normal.

And what will be the mean of y_i? Mean of $y_i = \beta_0 + \beta_1 x_i$ because you are assuming x i to be fixed and ε_i has mean 0. So, this will be here is the mean $\beta_0 + \beta_1 x_i$ and what will be the variance of y_i? That will be variance of β_0 which is constant so, 0 the variance of β_1 x_i which is constant so, 0. So, that will be same as the variance of ε_i which is σ^2.

So, I can write down here that y_i follows a normal distribution with mean $\beta_0 + \beta_1 x_i$ and variance σ^2 and if you and you can recall that we had express b_1 as $\sum_{i=1}^{n} k_i y_i$ right, I am not going into the details once again. So, based on that, I can write down from here that b_1 has got a normal distribution whose mean is β_1 and variance is σ^2 / s_{xx}.

Well, before I go further, let me clarify that I just finished the topic of maximum likelihood estimation and then, I started the test of hypothesis. So, please do not get confused that the test of hypothesis is based only for the maximum likelihood estimator right means that can be done for the ordinary least square estimation also and you have seen that at this moment, we are trying to construct the test of hypothesis and confidence interval for β_1 and β_0 which have got the same estimate in under the setup of least square estimation as well as maximum likelihood estimation.

(Refer Slide Time: 29:47)



Now, using this result, we can construct the statistics that $Z_1 = \frac{b_1 - \beta_{10}}{\sqrt{\frac{\sigma^2}{s_{xx}}}}$ and this will follow a normal distribution with mean 0 and variance 1 when H_0 is true, right.

There is a strong theory, I mean all those derivations we have in statistical inference, but we are not going to discuss those things here, you can just believe on me that this statistics will follow a N(0, 1) distribution under H_0. As soon as I say under H_0 that means, when H_0 is true.

So, this sentence and when H_0 is true, sometimes you will see that it is written in the book or sometime even I may use as under H_0. So, that is the same thing right. So, now the decision rule to accept or reject H_0 will be that reject H_0, the H_0 here is $\beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ at α level of significance if p value is less than α, ok.

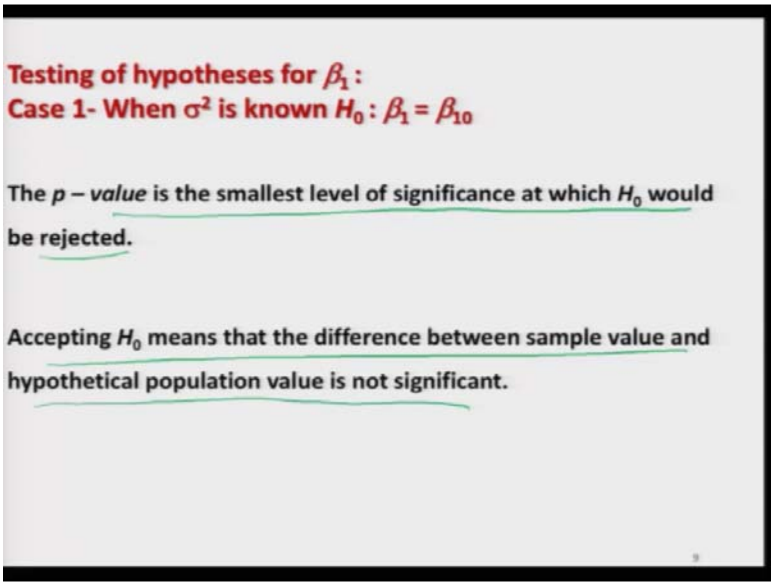
Means I can just give you one thing that when you try to conduct the test of hypothesis, then you have the sampling distribution of this statistics Z_1 which is say here N(0, 1) so, we try to fix here α so, it is a 2 sided in test of hypothesis because we are trying to take here not equal to. So, this region in the mid which is here the dotted region, this is the acceptance region and the shaded region here, this is the region of rejection.

What we try to do? We try to find out the critical value from the table of N(0, 1) which is here, this is the critical value and this is here the critical value and we try to see whether the calculated value of Z_1 which is here whether this lies in the region of acceptance or region of rejection and based on that, we try to take a call whether H_0 is going to be accepted or not.

So, this is the basic fundamental that is the theory, but since we are going to work in the software so, the software's usually give the outcome in terms of p value and then in that case, the rule is that reject H_0 when p value is smaller than α.

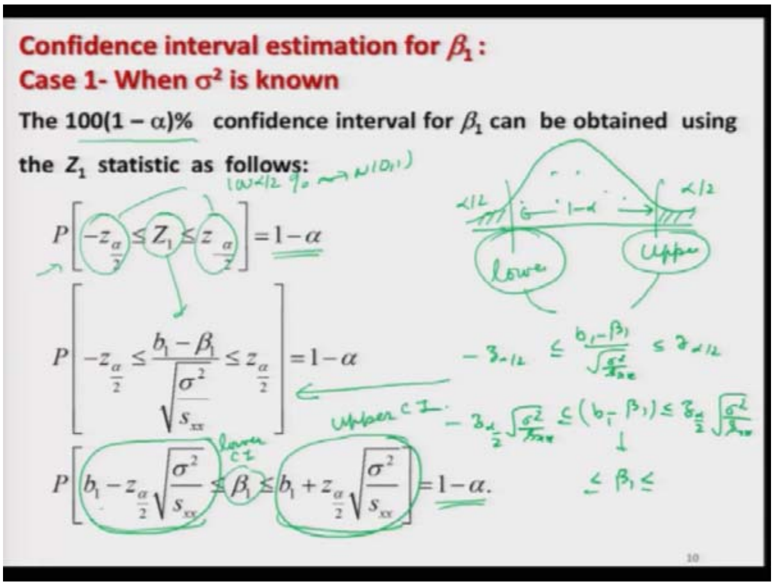
So, since we are going to do all the things on the R software so, I will not be considering the classical approach of test of hypothesis which is based on the tabulated values of either normal, t, chi square, f etc., but I will be depending more on the p value.

(Refer Slide Time: 32:50)



So, all the results whatever I am going to do here, they will be the based on p value. What is that p value? The p value is the smallest level of significance at which H_0 would be rejected and when I say that the H_0 is accepted, then accepting H_0 means that the difference between the sample value and the hypothetical population value is not significant right ok.

(Refer Slide Time: 33:16)



Now, I try to come to the construction of confidence interval under the same setup when σ^2 is known. So, we know in the case of confidence interval, what we try to do? We try to obtain here the lower limit and we try to obtain here the upper limit and suppose my confidence coefficient is $1 - \alpha$ so, usually when we are trying to consider a symmetric distribution like as N(0, 1) , then on the both side this shaded region will be $\alpha/2$ and this dotted region in the mid will be $1 - \alpha$, right.

So, these two values lower and upper limit, they are going to give us the values of the confidence interval. So, they are the lower and upper limits of the confidence interval. So, our objective is this, we want to find out these values based on the statistics. So, what we try to do here that we are assuming that my confidence coefficient is $1 - \alpha$ so, $100(1 - \alpha)\%$ confidence interval for β_1 can be obtained in this case, using the Z 1 statistics right.

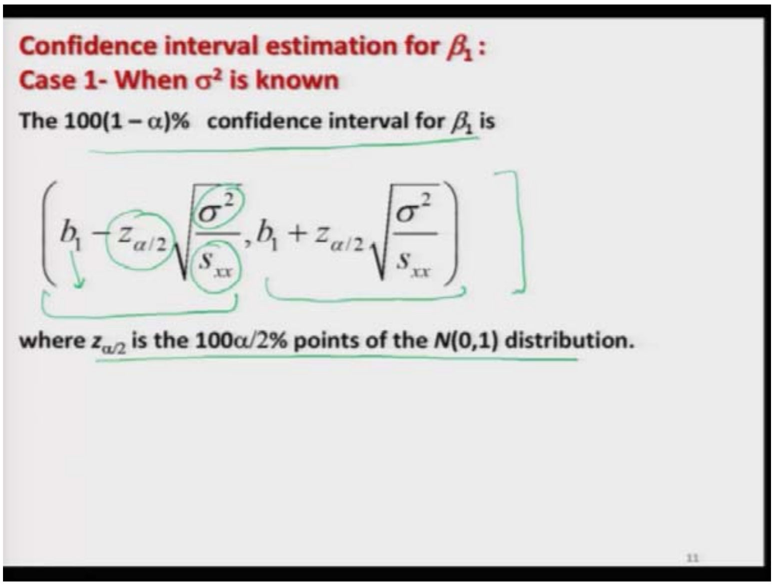
So, Z_1 is your you know nothing but $b_1 - \beta_1$ divided by standard deviation of b_1. So, what we assume that this capital Z_1 will be lying between $-z_{\alpha/2}$ and $+z_{\alpha/2}$. What are this $-z_{\alpha/2}$ and $+z_{\alpha/2}$? They are the say $100\alpha/2$ % points on the distribution of N(0, 1) right. So, what we try to do?

We simply write to this statement that the probability that the capital Z_1 will be lying between the critical values $-z_{\alpha/2}$ and $+z_{\alpha/2}$ and the probability of this event is $1 - \alpha$.

So, now, I can rewrite here it like this. Now, you can see here, I simply have to simplify it, you can write down this thing here b; $\dfrac{b_1 - \beta_{10}}{\sqrt{\dfrac{\sigma^2}{s_{xx}}}}$ which is lying between $-z_{\alpha/2}$ and $z_{\alpha/2}$ and this can be written here as say $b_1 - \beta_1$ lying between say $-z_{\alpha/2}$ and square root of σ^2/s_{xx} and here, $z_{\alpha/2}$ and square root of σ^2/s_{xx}.

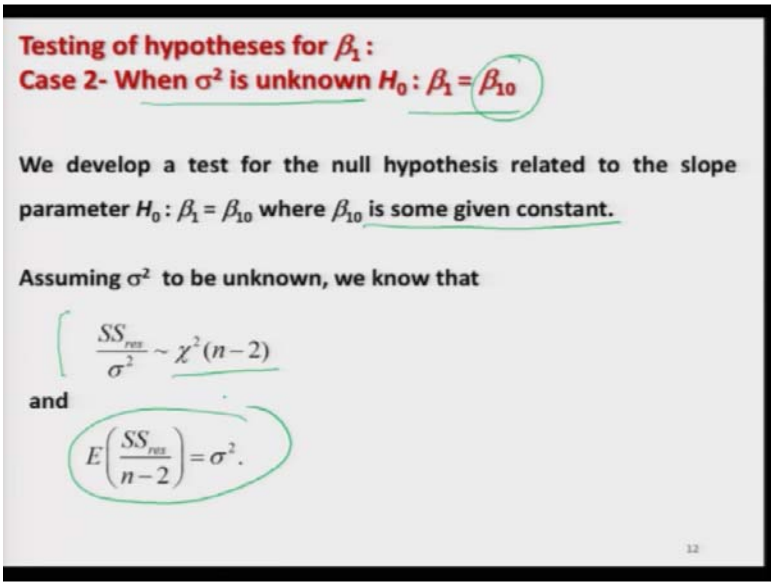
Now, if you try to solve it here, now this β_1 will be lying between these two limits which we have obtained here. So, that is the usual approach to obtain the confidence intervals. So, you can see here now, I have obtained the probability that β_1 will be lying between these two limits and these probabilities were $1 - \alpha$. So, this value here, this is the lower confidence limit and this value here, this is the upper confidence limit. So, lower confidence limit and the second one is upper confidence limit of the confidence interval right.

(Refer Slide Time: 36:32)



So, I can write down finally, that the 100(1 − α)% confidence interval for β 1 is given by these two limits where $z_{\alpha/2}$ is the 100α/2 % points of the N(0, 1) distribution. So, you can now see, this you have computed, you can obtain it from the software, this is known to us, this you can compute on the basis of given sample of data, this you can obtain from the table so, you can compute this entire limit similarly, you can compute the upper limit also and thus, you can compute the entire confidence interval. So, you can see here it is not difficult even if you want to compute it manually, but definitely, we are going to use it on the software.

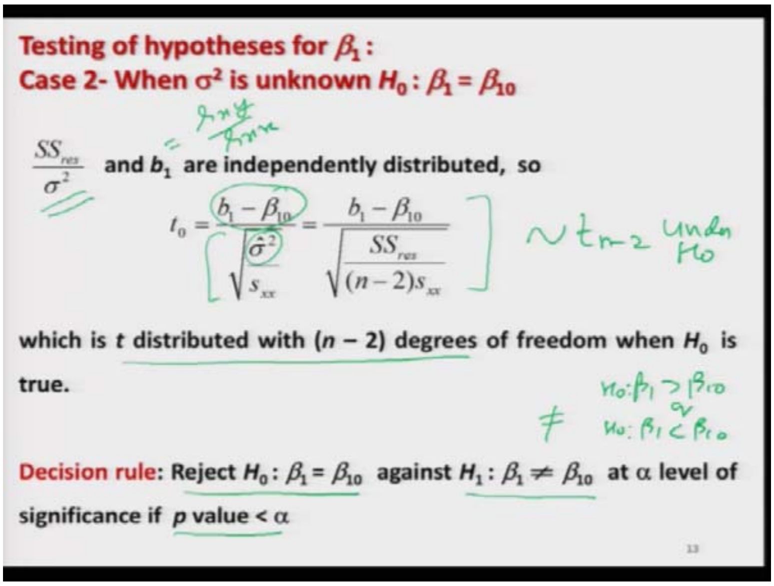
(Refer Slide Time: 37:20)



And now, I come to the 2nd case where we assume that σ^2 is unknown and we try to construct the hypothesis for the same parameter β_1, right. So, my hypothesis remain the same $H_0 : \beta_1 = \beta_{10}$ where β_{10} is some given value or given constant.

So, since we are assuming that σ^2 is unknown so, we have a result from statistical inference that SS_{res} upon σ^2 follows a chi square distribution with $n - 2$ degrees of freedom and if you remember, we had already used it in the earlier lecture when we were trying to estimate the value of σ^2 based on the least square estimation and we had obtained there that expected value of $SS_{res}/(n - 2)$ is σ^2 right.

(Refer Slide Time: 38:12)



So, now, actually I can also prove that if you try to see here that this is your here ss_{res} and b_1 was your s_{xy}/s_{xx}. These two quantities are independently distributed although, I am not proving it here, but that is pretty simple and right.

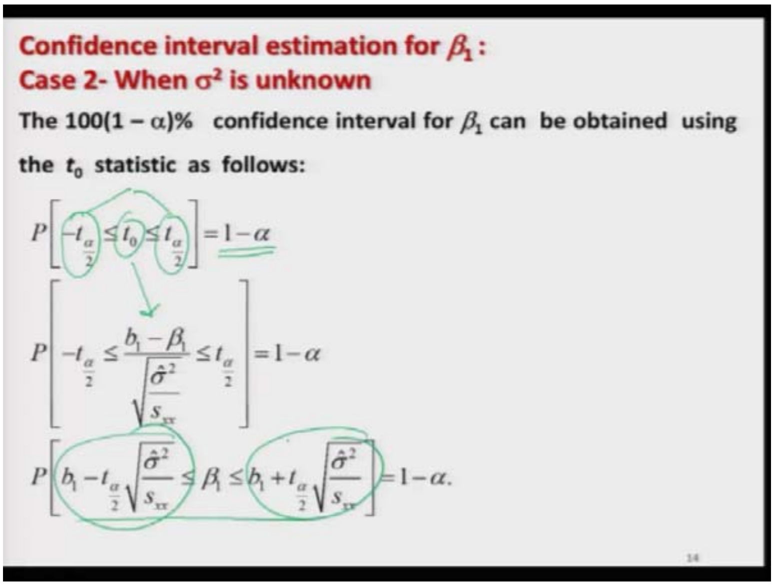
So, and if you can show that these two are independently distributed which are actually independently distributed, you can try, then $t_0 = \frac{b_1 - \beta_{10}}{\sqrt{\frac{\hat{\sigma}^2}{s_{xx}}}} = \frac{b_1 - \beta_{10}}{\sqrt{\frac{SS_{res}}{(n-2)s_{xx}}}}$.

And if you try to write down this statistic, this will follow a t distribution with n – 2 degrees of freedom when H_0 is true right. So, this simply follows $t_{(n-2)}$ under H_0 that is when H_0 is true. So, now, I can frame the decision rule which is pretty simple that reject $H_0 : \beta_1$ equal to β_{10} against $H_1 : \beta_1$ is not equal to β_{10} at α level of significance if p value is less than α ok.

Before I go further, let me try to clarify one point. In this test of hypothesis, in this chapter, we are mostly interested in it testing the alternative hypothesis which is of the form of not equal to that is two-sided alternative hypothesis. Actually, this test of hypothesis is going to play a very important role when in the variables collection and in knowing whether a variable is important or good or bad right.

So, that is why we are considering here only the two-sided test of hypothesis otherwise, the left tilt test or right tilt test something like β_1 is greater than β_{10} or β_1 is less than β_{10}, these type of hypothesis can easily be done exactly on the same line using the same statistics. So, do not get confused that why I am not considering here other type of test of hypothesis, they can be done exactly on the same way.

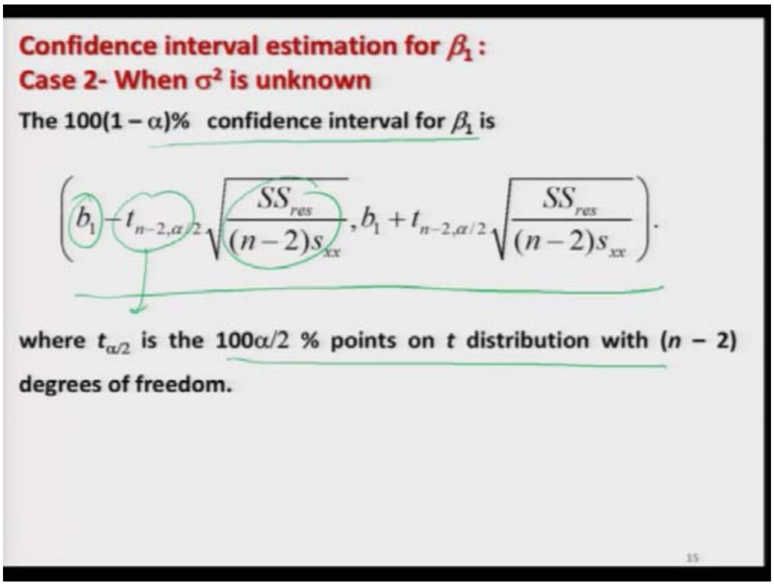
(Refer Slide Time: 40:42)



And similarly, if you want to construct the confidence interval, exactly you have to follow the same thing. Now, I will assume the $100(1-\alpha)\%$ confidence interval for β_1 can be obtained using the t_0 statistics and we assume that this t_0 lies between $-t_{\alpha/2}$ and $+t_{\alpha/2}$ where these two values are the critical values which are obtained from the t tables that is the table for the t probabilities and the probability of such an event is $1-\alpha$.

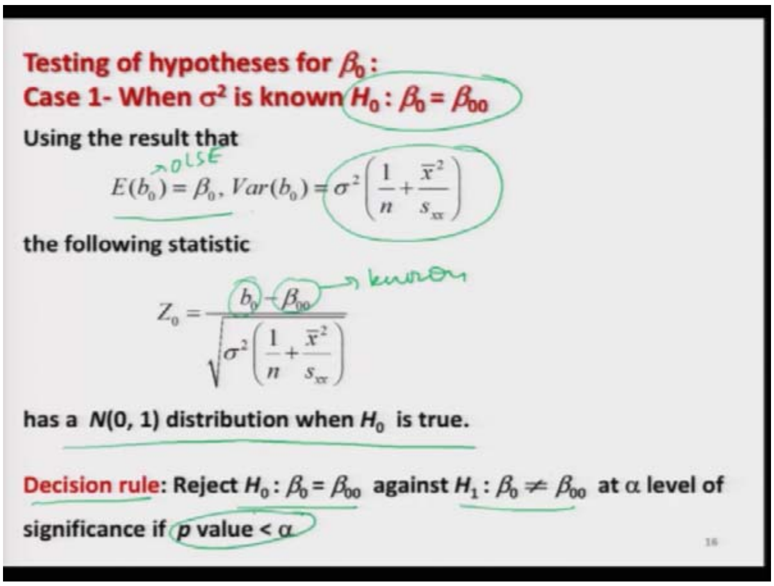
So, you simply try to substitute here the value of t_0 and just try to simplify the simplify it exactly and the same way as we have done here right. Exactly, in the same way as we have done here, try to simplify it, very simple and then, you will obtain here this inequality and the probability that β_1 is lying between these two limits is equal to $1-\alpha$.

(Refer Slide Time: 41:43)



And hence, I can obtain the $100(1-\alpha)\%$ confidence interval for β_1 when σ^2 is unknown as here like this. So, you can see here all these things can be obtained from the data right and about this here t value that can be obtained from the table, but anyway I am not going to show it here manually because we are going to use the software. So, here, what is this thing? This $t_{\alpha/2}$ is the $100\alpha/2$ % points on the t distribution with $n-2$ degrees of freedom ok.

(Refer Slide Time: 42:18)



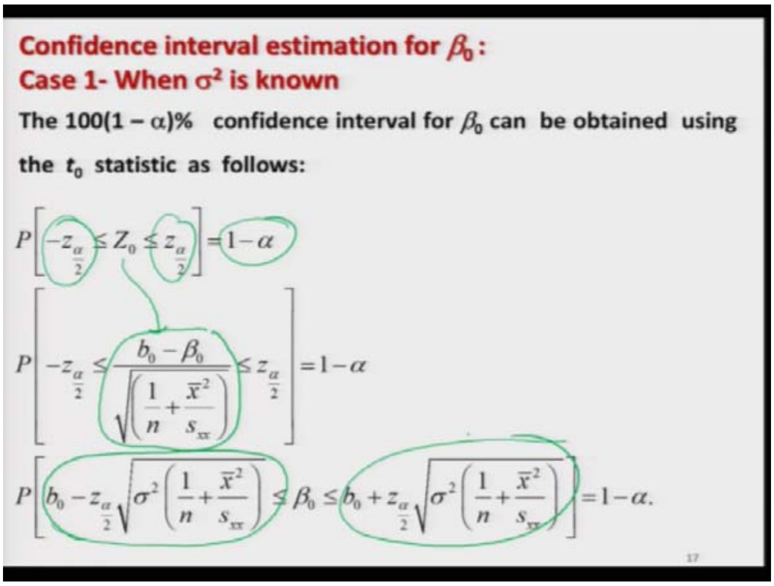
Now, I come to the test of hypothesis for the β_0. Now, the test of hypothesis and confidence interval for β_0 under the case of σ^2 and σ^2 when σ^2 is known and when σ^2 is unknown can be constructed and develop exactly on the same line as I have done for β_1. The only thing is this, the variance of β_0 and estimate of variance of $\hat{\beta}_0$ which is the variance of b_0 and $\widehat{\text{var}(\hat{b}_0)}$, they will be only changed, right.

So, under this case, you can see here we already have proved that expected value of b_0 which is the ordinary least square estimator of β_0, this is an unbiased estimator of β_0 and the variance of b_0 can be obtained by this expression right that we already have obtained in the earlier lecture.

So, now, I can use the Z statistics and I can construct my statistics by $b_0 - \beta_{00}$ which is a known quantity here, which is a given quantity here because my hypothesis is here H_0 $\beta_0 = \beta_{00}$ or β_{00} or β_{00} you can say.

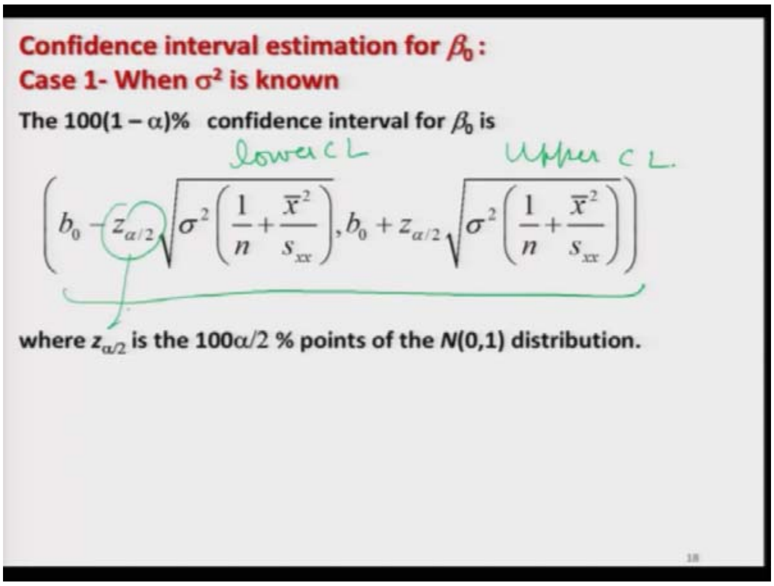
So, this will also have a $N(0, 1)$ distribution when H_0 is true. So, you can write your here decision rule exactly on the same line that you reject this H_0, against H_1 $\beta_0 \neq \beta_{00}$ at α level of significance if p value is smaller than α.

(Refer Slide Time: 44:01)



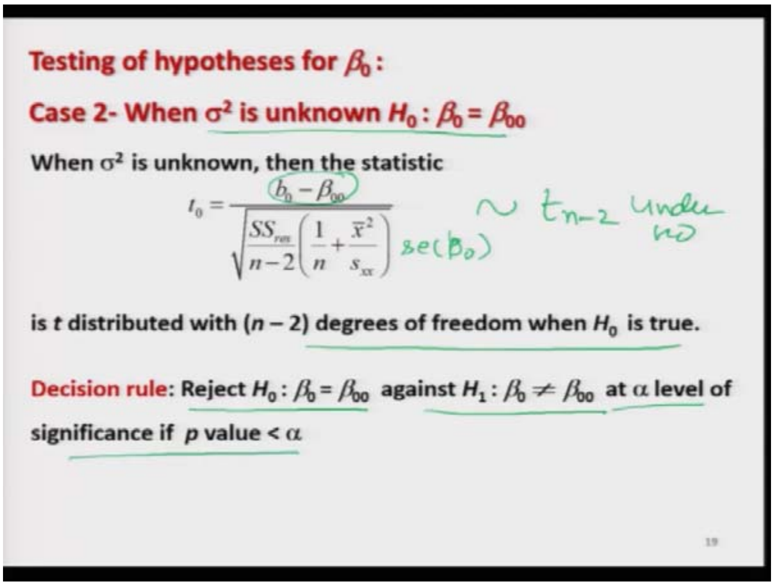
And if you want to construct the confidence interval, you have to use the same technique that Z_0 which is given here like this is lying between $-z_{\alpha/2}$ and $+z_{\alpha/2}$ and the probability of such an event is $1 - \alpha$, right. So, we can simply solve this equation as we have done earlier and two times so, and you will obtain the lower limit here by this expression and upper limit by this expression.

(Refer Slide Time: 44:34)



And so, I can say that the $100(1-\alpha)\%$ confidence interval for this β_0 is given by these two limits, this is the lower confidence limit and this is the upper confidence limit right and as usual, this $z_{\alpha/2}$ is the $100\alpha/2$ % points on the $N(0, 1)$ distribution.

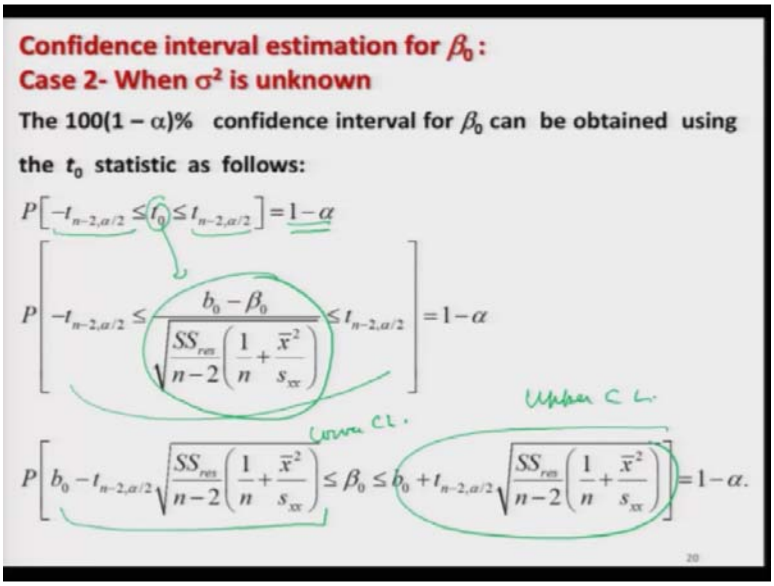
(Refer Slide Time: 44:57)



And similarly, when σ^2 is unknown, you know now and this case, we will have a t statistics and this statistics is given by $b_0 - \beta_{00}$ divided by the standard error of, this is the standard error of b_0 and if you try to find out using the rules of probability theory and probability distribution, this statistics will have a t distribution with $n - 2$ degrees of freedom under H_0 or when H_0 is true.

So, now, I can construct the decision rule that reject H_0 again this two-sided H_1 at α level of significance if p value is smaller than α. So, now you can see that now means everything is very very similar once you have done the first case doing all other cases have the same story means everything is on the similar lines.

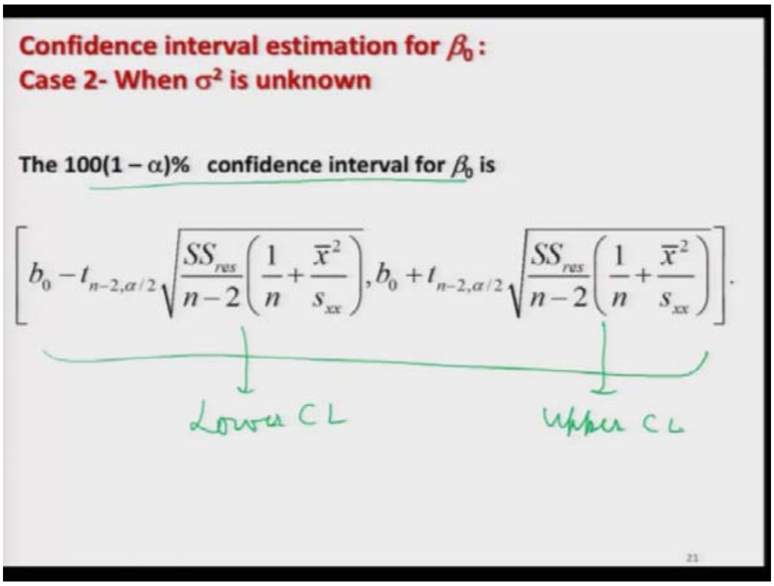
(Refer Slide Time: 45:57)



And similarly, you can also construct the confidence interval. So, in this case, the t_0 which is here this statistic will lie between $-t_{\alpha/2}$ and $+t_{\alpha/2}$ at $n-2$ degrees of freedom and the probability of such an event is $1 - \alpha$.

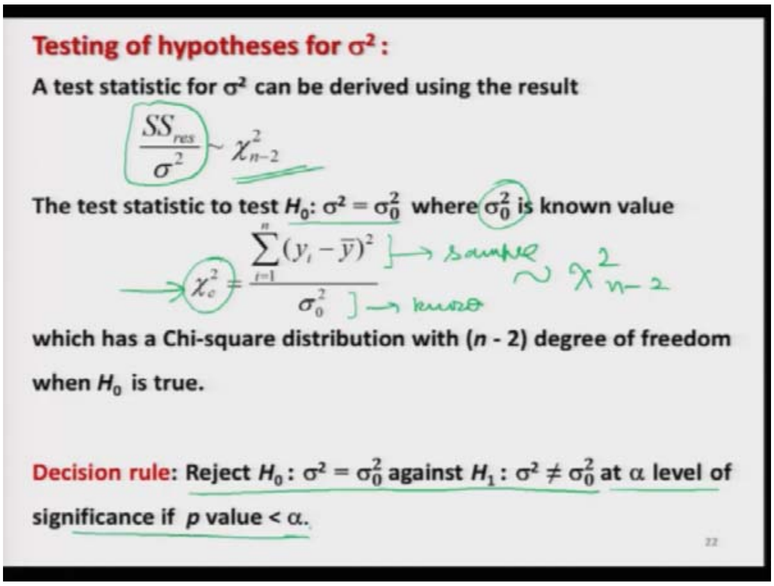
So, if you simply try to solve this inequality, you will get here this as your lower confidence limit and this quantity here as a upper confidence limit and the probability of such an event will be $1 - \alpha$. So, the confidence coefficient is $1 - \alpha$.

(Refer Slide Time: 46:30)



Hence, the $100(1-\alpha)\%$ confidence interval for β_0 can be obtained here like this right. So, now, we have obtained the confidence interval and test of hypothesis for β_0 and β_1 under two cases when σ^2 is known and σ^2 is unknown. So, this is your here lower confidence limit, and this is your here upper confidence limit.

(Refer Slide Time: 47:01)

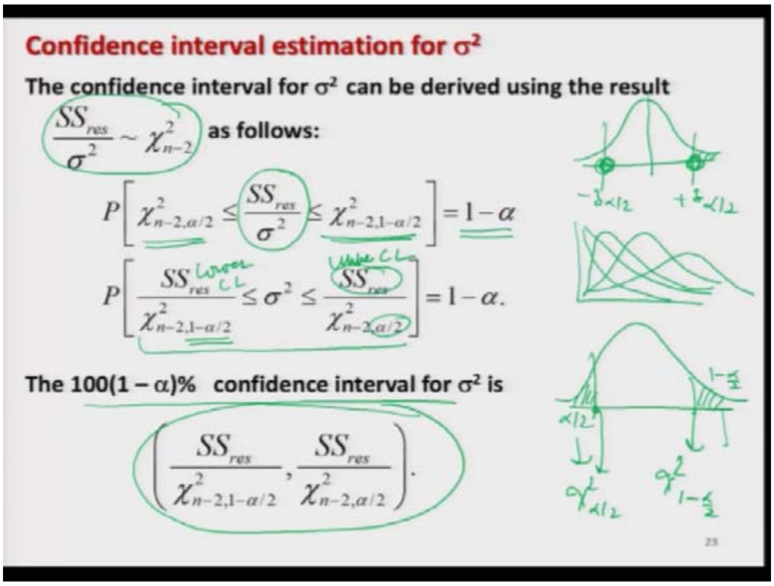


So, after this we consider the test of hypothesis and confidence interval for σ^2. So, the test statistics for σ^2 is found using the result that sum of square due to residual divided by σ^2 follows a chi square distribution with $n-2$ degrees of freedom.

So, based on this, the test statistics for the hypothesis $H_0 : \sigma^2 = \sigma_0^2$ where σ_0^2 is some known, value given value constant can be constructed and the final statistic comes out to be like this $\chi_c^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma_0^2}$. So, this value can be obtained from the sample and this value is known. So, whole this chi square c statistics can be found manually also.

And this statistics will follow a chi square distribution with $n-2$ degrees of freedom under H_0 that is when H_0 is true ok. So, based on this, the decision rule can be made as a reject H_0 against H_1, two-sided hypothesis at α level of significance if p value is smaller than α, same thing what we have done earlier.

(Refer Slide Time: 48:23)



And you can also construct the confidence interval using the same result, but now, here you have to be careful normal and t distribution they are symmetric like this one, symmetric around their mean value right, but for chi square, chi square is not a symmetric distribution, and it has different shapes depending on the degrees of freedom like this right. Depending on the degrees of freedom is less 1, 2, 3, 4, then even the curve is like this.

So, in this case, you cannot have a structure like that $\alpha/2$ on the same inside will have the same value on the x-axis like we have done earlier that $-z_{\alpha/2}$, then this is going to be $+z_{\alpha/2}$, this will not hold, but in this case, what we are going to do? Suppose this is here see here $\alpha/2$, then this is here $1 - \alpha/2$.

So, you have to actually look into the book that how they are trying to consider. Sometime they take this area on the left-hand side to $b_1 - \alpha/2$ and right hand side area to be $\alpha/2$ so, you have to just check before you read a book. So, in this case, we are simply trying to say here suppose this is my here chi square value with $\alpha/2$ and this is here the value which is chi square $1 - \alpha/2$.

So, now I can say here that this statistics SS_{res} upon σ^2 will lie between these two limits on the chi square distribution and the probability of such an event will be $1 - \alpha$. And if you simply try to simplify this inequality, you get here these two limits.

So, this is going to be the lower confidence limit, and this is going to be the upper confidence limit and this will come out to be SS_{res} divided by the corresponding value of chi square at $1 - \alpha/2$ level significance level and SS_{res} divided by chi square value at $\alpha/2$ significance level.

And based on that, the $100(1 - \alpha)\%$ confidence interval will come out to be like this. So, you can see here that this is a function of SS_{res} that is sum of square due to residuals and here, you have to be careful that and always remember that chi square is not a symmetric distribution.

So, now, we come to an end of this lecture. Although, pretty long lecture, but my objective was to finish in a single shot so that your time is saved and I do not have to repeat the things between two lectures. I hope you do not mind it. Well, you can means listen to it in two parts.

Well, so, my objective was here that I have given you the main idea about the test of hypothesis, your question will be where it is going to be used, but believe me for some time, this is one of the very important tool in statistical modeling, particularly under the linear regression analysis and I will show you later on that without this, you just cannot work in linear regression analysis.

Yes, sometime people, students get confused that why I am trying to do all this theory, but when you come to multiple linear regression model, I am promising you each and everything whatever you have done that will play a very important role and without understanding these things, you cannot do it and my problem is this here, I can show you all the things, I can create a confidence interval.

I can give you all the values individually, but in case of multiple linear regression model, I will not be able to show you things more clearly, but you will have to assume, you will have to think in your mind that this is how the things will look like that is the idea right. So, have patience, have confidence on me, have faith on me and try to revise it, try to see it, try to settle down this concept in your mind and I will see you in the next lecture till then, goodbye.