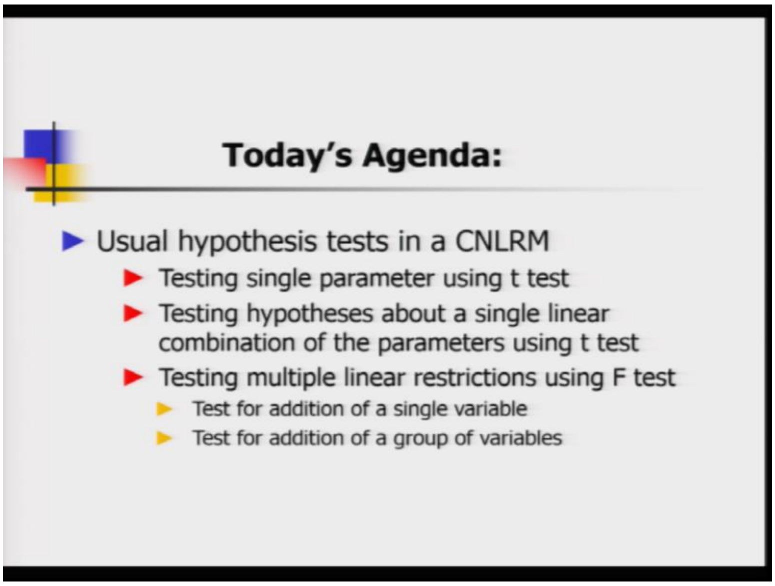
Applied Statistics and Econometrics
Professor. Deep Muhkerjee
Department of Economic Sciences
Indian Institute of Technology, Kanpur
Lecture No. 28
Hypothesis Testing with CNLRM

Hello students, welcome back to the lecture series on Applied Statistics and Econometrics. So, today we are going to revisit the concept of hypothesis testing but we are going to discuss hypothesis testing in the context of linear regression analysis. So, before we go back to the old friends like t test, f test, et cetera, let us have look at today's agenda items.

(Refer Slide Time: 0:38)



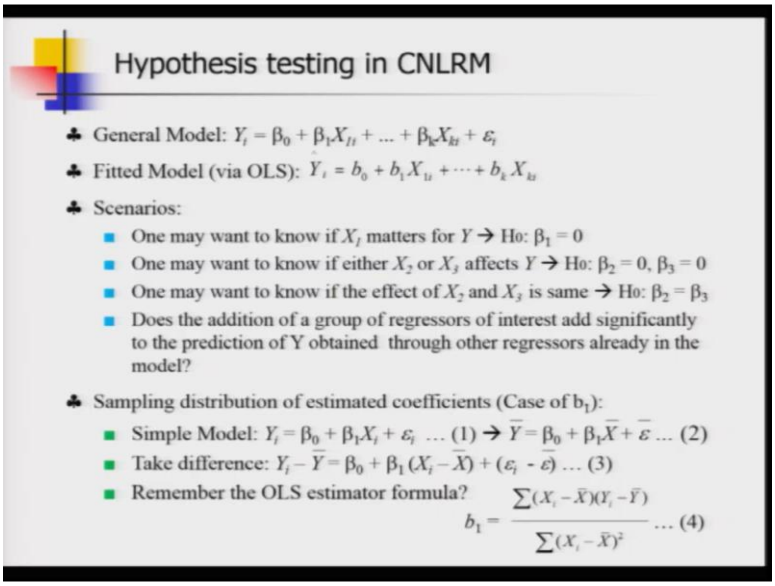
Today's Agenda:

- Usual hypothesis tests in a CNLRM
  - Testing single parameter using t test
  - Testing hypotheses about a single linear combination of the parameters using t test
  - Testing multiple linear restrictions using F test
    - Test for addition of a single variable
    - Test for addition of a group of variables

So, today we are going to study the hypothesis testing in the context of classical normal linear regression model. And of course, hypothesis testing in the context of CNLRM is a vast area so we cannot finish the discussion in one lecture. So, in today's lecture I am going to focus on 3 major items and they are testing single parameter value using t test and then I am going to talk about another use of t test while we want to test hypothesis about a single linear combination of the parameters.

And finally we are going to talk about testing multiple linear restrictions and there we are going to see the use of f test. So, here throughout this lecture we are going to consider a multiple linear regression model, it implies that we are going to work with k number of explanatory variables. So, let us have a look at the general model and then from there we will develop hypothesis testing.

(Refer Slide Time: 1:43)



Hypothesis testing in CNLRM

- General Model: $Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \varepsilon_i$
- Fitted Model (via OLS): $\hat{Y}_i = b_0 + b_1 X_{1i} + \cdots + b_k X_{ki}$
- Scenarios:
  - One may want to know if X_1 matters for $Y \rightarrow$ Ho: $\beta_1 = 0$
  - One may want to know if either X_2 or X_3 affects $Y \rightarrow$ Ho: $\beta_2 = 0, \beta_3 = 0$
  - One may want to know if the effect of X_2 and X_3 is same $\rightarrow$ Ho: $\beta_2 = \beta_3$
  - Does the addition of a group of regressors of interest add significantly to the prediction of Y obtained through other regressors already in the model?
- Sampling distribution of estimated coefficients (Case of b_1):
  - Simple Model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ … (1) $\rightarrow \overline{Y} = \beta_0 + \beta_1 \overline{X} + \overline{\varepsilon}$ … (2)
  - Take difference: $Y_i - \overline{Y} = \beta_0 + \beta_1 (X_i - \overline{X}) + (\varepsilon_i - \overline{\varepsilon})$ … (3)
  - Remember the OLS estimator formula?

$$b_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \text{ … (4)}$$

So, our general model has this equation, Yi equals to beta naught plus beta 1 X1i dot dot dot plus epsilon i. So, here we have k number of explanatory variables and the associated regression coefficients are beta 1 to beta k, beta naught is my intercept parameter and epsilon i is the normal error. Or you can say that this is stochastic disturbance term which follows normal distribution. So, that is basically the mother model of CNLRM.

Now we fit regression equation via OLS method and we get Y hat i, that is the fitted value of my dependant variable y. And I get the estimators of the unknown population parameters and their b naught, b1 dot dot dot bk. So, now let us discuss different scenarios where hypothesis testing could be very handy. Now, one may want to know whether an explanatory variable in particular say X1 matters for y. So here I want to say that whether X1 actually helps in explaining any bit of variation in y or we can also say that whether X1 determines the values of y.

So, that could be analysed by setting a null hypothesis beta 1 equals to 0, because if beta 1 that unknown population parameter value is indeed 0, then that implies that there is no linear relationship between the explanatory variable X1 and the dependant variable y. Similarly, one may want to know if either X2 or X3 affects y. So, here you note that I am talking about a simultaneous occurrence of two conditions, beta 2 equal to 0 and beta 3 equal to 0. So, it could be the case that beta 2 is equal to 0 but beta 3 is not equal to 0, and beta 3 could be 0 and beta 2 could not be 0.

So, there are different combinations possible. So, this is basically the case of the joint hypothesis. And one may also want to know if the effect of X2 and X3 is same. So, in that case we set a null hypothesis that will say that beta 2 equal to beta 3. Now, apparently it may look very silly that why we want to have that kind of hypothesis testing, but that is very useful especially when you have a small dataset and you want to save degrees of freedom.

So, if you indeed can prove that beta 2 is equal to beta 3 or the regression coefficients for two explanatory variables are indeed the same, then you can use the sum of these two variable, then we can use the sum of two variables X2 and X3 as the explanatory variable and then by doing that we save one degrees of freedom. Because there is only one common slope coefficient to be estimated in this case.

Now one can also ask a very interesting question, does the addition of a group of regressors of interest add significantly to the prediction of y obtained through the other regressors already in the model? So, basically what do I mean to say? If I remind you about the discussion from last lecture about adjusted R square and all, then you probably can understand what I am trying to say here.

As you keep on adding more and more explanatory variables in the regression equation, your R square will go up but not necessarily adjacent R square will go up. And this is actually looking at the same problem from the hypothesis testing point of view. For adjusted R square you cannot have a hypothesis testing, it is just a summary measure, it is a model fit measure.

But you can make use of hypothesis testing procedures to figure out whether you should add a couple of more variables in the existing set of regressors. If you do not find that this group of additional regressors explain significant amount of extra variation in y, then you should not add these extra variables in the model. So, actually hypothesis testing will also help you in model selection.

Now, before we go to the formal t test and f test m all, we have to look at the sampling distribution of the estimated coefficients. Now, why we have to look at the sampling distribution of the coefficients?

Note that here, unlike the previous time when we had one single variable and we were interested in either comparing the mean or the variance across two groups or two populations, here actually we are looking at a different kind of problem. So, here we have a linear regression setup where we have multiple regressors, explanatory variables and they are supposedly linearly related with the dependant variable y.

So, once you have one sample, you can apply the OLS technique and you can get one set of b naught, b1, b2, et cetera. And when you get another sample from the same population, then you can get another set of values b naught, b1, b2, et cetera. So, as you keep on drawing sample from the same population again and again, you are going to generate different sets of values for b naught, b1, b2, et cetera.

So, then basically you are going to create a sampling distribution for the regression coefficients, estimated regression coefficients. So, that is why the b1, b2, b naught, et cetera, they will have distribution of their own. So, of course, if they have the distribution of their own, then there will be a mean and a variance, so we can see from the previous lectures, the Guss Markov theorem and all that the mean of the sampling distribution is basically equal to the unknown population parameter value. So, one can test a particular value of the population parameter by looking at the mean of this sampling distribution.

So, here to explain the situation, let us come down to two variable model where we have one single x, it will help us to simplify the situation, otherwise I have to introduce the matrix algebra again and it will be very clumsy and probably we will not get any extra mileage by introducing that clumsy matrix algebra. So, for that purpose, one single explanatory variable is good enough.

So, we have a simple model and let us now rewrite this simple model equation, say equation 1. So, what we are going to do, we are going to take the mean of the left hand side variable and the right hand side expression. So, here you know we get Y bar equals to beta naught plus beta 1 times X bar plus epsilon bar, and that is m equation 2. So, it simple, basically if you have n number of observations, you first apply sum on both sides of the equation and then you divide by the number of observations in and that is the way you get the sample mean for Y and X.

So, once you have the sample mean, then you take the difference and that will give you a new equation 3 and that is Yi minus Y bar equals to beta naught plus beta 1 times Xi minus X bar plus epsilon i minus epsilon bar. So, you see here in equation 3, I have expressed my variables Y X and epsilon in the deviation form. Now, why have I done so? You will see very soon why I have done this. There is a reason behind it.

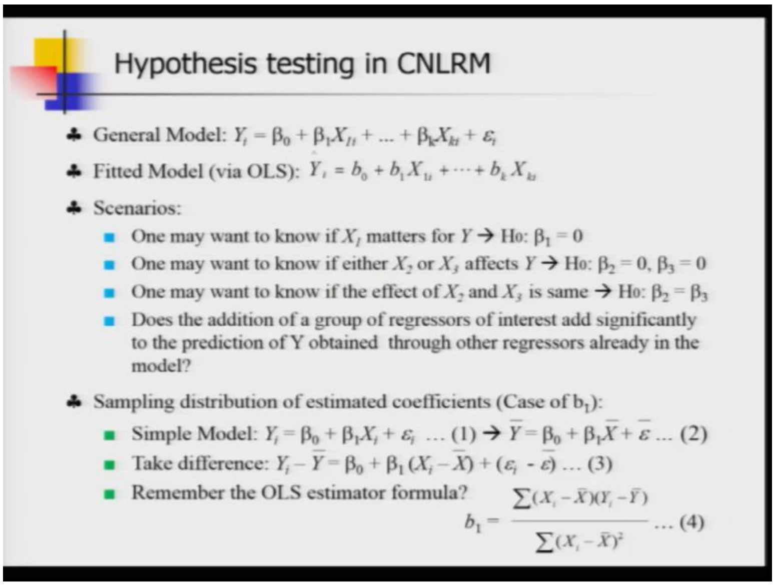
So, now I would like to draw your attention to the OLS estimator formula. So, in this simple linear regression model, you are going to get two estimator formula, one for your intercept parameter and one for the slope parameter. So, now in economics as I told you that we are mostly interested in the slope parameters, so here I am only going to show you the case of beta 1. So, the OLS estimator for beta 1 is given by b1 and once you know the value of b1, you can actually find out the value of b naught also which is the estimator for unknown population parameter beta naught.

But forget about that, let us concentrate on equation number 4. So, here I am showing you the OLS estimator formula for b1. Now you concentrate on the numerator and in the denominator. So, here in the numerator, you have the multiplication of two variables which are in deviation form, right. Now you go back to equation 3 and you figure out that there is Yi minus Y bar expression there. So, you can replace that expression for Yi minus Y bar in equation 4 by what you see in equation 3.

So, now you have in equation 4 everything in terms of two variables, so the deviation is for two variables, one is the explanatory variable X and one is the random stochastic disturbance term epsilon. Now you note that X is fixed, X does not have any distribution of its own, but epsilon we are assuming that it is a random variable, so it has its own distribution and we are assuming here the case of normality, so we are assuming that epsilon, the stochastic disturbance term actually follows normal distribution.

So, now I am going to make use of some interesting theoretical properties from normal distribution and that will help us in doing hypothesis testing. So, in the next slide we are going to see this. So, note down one thing before we leave this particular slide that now my b1, the estimator formula actually becomes a linear combination of this random variable epsilon i. So, that is very critical point.

(Refer Slide Time: 12:36)



Hypothesis testing in CNLRM

- General Model: $Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$
- Fitted Model (via OLS): $\hat{Y}_i = b_0 + b_1 X_{1i} + \cdots + b_k X_{ki}$
- Scenarios:
  - One may want to know if X_1 matters for Y → Ho: $\beta_1 = 0$
  - One may want to know if either X_2 or X_3 affects Y → Ho: $\beta_2 = 0$, $\beta_3 = 0$
  - One may want to know if the effect of X_2 and X_3 is same → Ho: $\beta_2 = \beta_3$
  - Does the addition of a group of regressors of interest add significantly to the prediction of Y obtained through other regressors already in the model?
- Sampling distribution of estimated coefficients (Case of b_1):
  - Simple Model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ … (1) → $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{\varepsilon}$ … (2)
  - Take difference: $Y_i - \bar{Y} = \beta_0 + \beta_1 (X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon})$ … (3)
  - Remember the OLS estimator formula?

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \ldots (4)$$

Now we start from that point, and as the errors are normally distributed as per our normality assumption, we can say that the linear combination of errors is also normally distributed. So, in the last slide only I have shown you that the OLS estimator b1 actually is a linear combination of the random variable epsilon. So, the estimator formula b1 will also be normally distributed, because it is a linear combination of normally distributed variables. So, focus on this particular result where I am saying that b1 follows a normal distribution with mean beta 1 and variance sigma square divided by sum of Xi minus X bar square.

So, I have shown you in the previous lecture that Guss Markov theorem tells us that the OLS estimator are blue, the base linear unbiased estimators. And their unbiasedness means that the expected value of the estimator will give me the value of the true population parameter. So, expected value of b1 will give me beta 1, that specifically in a nutshell Gauss Markov theorem tells us. And here, we are going to make use of that theorem, so now we know that as OLS estimators are unbiased so b1 will have a mean of beta 1.

Now, from where did I get the value of the variance? It requires a lot of complicated calculations and we are skipping this because this is not a theoretical course, so I am only going to make use of that theoretical result. Now note that the variance of the sampling distribution for b1 is basically cannot be computed. Why? Because sigma square is unknown. So, instead of sigma square we can make use of the estimate, sample estimate X square, the sample variance.

Do you remember the formula of root mean square errors? So, if you remember when we discussed the classical linear regression model in first two lectures, there we said that once the regression residuals are obtained by fitting the curve and then getting the fitted values from the fitted lines, then you can make use of those residuals to calculate the standard deviation of these residual vector and that basically gives the root mean square error. And needless to say that you have to take care of the degrees of freedom because you are going to estimate the slop and intercept parameters of the fitted lines before you actually start the calculations for root mean square errors.

So, now here that concept, root mean square error that we have learned previously can be used here. So, now sigma can be replaced or we can use proxy for sigma by utilizing the quantity s that we have calculated from our data set. So, if you have forgotten the formula root mean square error, I suggest that you go back to the previous lecture. Now in a nutshell that is the square root of the mean of the square of all the error terms or residual terms.

So, now by making use of the root mean square error, I can actually write the standard deviation of the estimated slope coefficient and the standard deviation of the sampling distribution of one regression coefficient is called standard error. So, here I am showing you the standard error formula for b1 slope coefficient and that is equal to S, that is basically the root means square error divided by square root of sum of Xi minus X bar square.

Now we are going to start our journey with hypothesis testing. So, first we are going to study the case of a single parameter. So, once you run a regression you may be interested to know that whether a particular variable say Xj is playing good enough role in explaining the variation of Y. Or in other words, whether Xj has a determining role to predict the future values or out of sample values for Y. So, for that actually we have to contact a t test.

So, now let us see how we can bring that old friend back in this linear regression context. So, here as we have got the standard error or standard deviation of the b1 slope coefficient, now we are in a good shape, we can go for the test statistics formation and that is denoted by t obs or t observed and that is actually given by b1 divided by standard error of b1. Now why I have minus 0 here?

Because note, here when we are conducting t test for a single parameter in a regression equation, the null hypothesis says that the particular regression coefficient, unknown population parameter takes value 0 and the alternative hypothesis is that no it is not 0. Now we are not saying that whether it is alternatively positive or negative, so we are going for a two tail test and we are just setting the alternative hypothesis saying that beta j is not equal to 0 and of course, the null is beta j is equal to 0.

So, it is taking a particular value of that unknown population parameter. And if you remember, we have to deduct that particular parameter value that is given in the null hypothesis while we are constructing the t statistic and that is why I am deducting 0 from b1, because 0 is the hypothesized value of beta 1. So, we have to deduct the hypothesized value of the unknown population parameters from the observed value of the sample statistic. So, following that principle from hypothesis testing, we are deriving or defining our t observed like this.

So, now these test statistic follows a t distribution with capital N minus K, capital K degrees of freedom if error is normally distributed. And we can see that in the small sample case it is better to assume that my test statistic t observed follows a t distribution with degrees of freedom and all, but when you have a large sample, say, when number of observation is greater than 40 or so, but we can also make use of the normal distribution in conducting this kind of t test.

So, if we have capital N minus capital K value is greater than 30, in that case, we can assume normality and go ahead with the standard normal table to find the critical value, but otherwise you consult a table to find the critical value. Okay, so one can assume that the t statistic will follow a standard normal distribution when N tends to infinity and in that case you feel free to consult the standard normal table.

Now we have to choose one particular level of alpha and let us choose a particular alpha, norm is 5 percent, if you remember, and now we have to declare our rejection rules. So, there are 2 rejection rules depending on the approach you are taking. So, if you are taking the classical or traditional approach, then you have to compared the observed value of test statistic with the critical value that you found from the statistical table.

So, you reject null hypothesis if the absolute value of the observed test statistic is greater than the absolute value of the critical value that you find from the statistical table. And if you follow the p value method, then basically you reject null hypothesis if p value is less than the chosen alpha level. So, generally alpha is 0.05 in maximum of the imperial work, so if p value is less than 0.05, then you reject the null hypothesis.

And in this slide we are going to talk about the confidence interval and confidence interval for the regression coefficients can be formed by following the simple formula here and that is basically estimated coefficient plus minus the critical t value multiplied by standard error of coefficient.

Now why do we have to bring that concept of confidence interval in these case? It is because if you remember our previous lessons from hypothesis testing discussions, there I told you that you can also comment on the significance of a particular regression coefficient by looking at the confidence interval.

So, in the context of regression, here suppose you make a confidence interval for one beta coefficient and then you see that 0 is included in that confidence interval. Then you cannot rule out the 0 value for the population parameter and then you can actually judge that, okay, this particular variable may not have any effect in determining variation of Y, so the effect of X is insignificant. Or you can say that the coefficient that is associated with the explanatory variable is insignificant. So, confidence interval can also be very useful to check your results that you are obtaining from the hypothesis testing exercise.

(Refer Slide Time: 23:15)

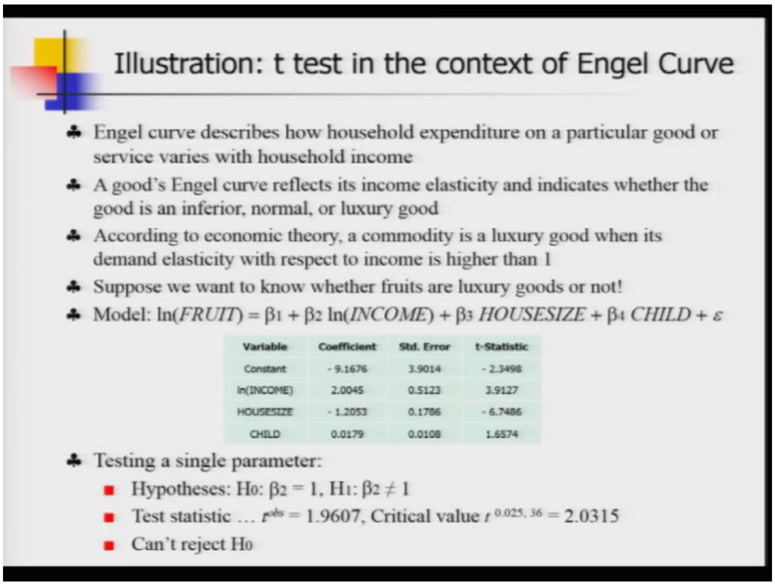


Illustration: t test in the context of Engel Curve

- Engel curve describes how household expenditure on a particular good or service varies with household income
- A good's Engel curve reflects its income elasticity and indicates whether the good is an inferior, normal, or luxury good
- According to economic theory, a commodity is a luxury good when its demand elasticity with respect to income is higher than 1
- Suppose we want to know whether fruits are luxury goods or not!
- Model: $\ln(FRUIT) = \beta_1 + \beta_2 \ln(INCOME) + \beta_3\, HOUSESIZE + \beta_4\, CHILD + \varepsilon$

Variable	Coefficient	Std. Error	t-Statistic
Constant	- 9.1676	3.9014	- 2.3498
ln(INCOME)	2.0045	0.5123	3.9127
HOUSESIZE	- 1.2053	0.1786	- 6.7486
CHILD	0.0179	0.0108	1.6574

- Testing a single parameter:
  - Hypotheses: H_0: $\beta_2 = 1$, H_1: $\beta_2 \neq 1$
  - Test statistic … $t^{obs} = 1.9607$, Critical value $t^{0.025,\,36} = 2.0315$
  - Can't reject H_0

Now, I am going to show you an illustration of the t test and we bring an example from microeconomic theory. So, in the last lecture only we have spoken about Engel Curves. So, now it is a good time to revisit that concept again and discuss case of t test. So, let me now remind you again so if you know jog your memory, probably you will remember that Engle Curves describe how household expenditure on a particular good or service varies with household income.

And the shape of Engel Curve can be of various form, it could be highly non-linear, it could be linear in log plane. So we are not going to comment on a particular shape of Engel Curve, although we know we discuss that on particular function for may be very useful in fitting Engel curves in real life data, but here we are interested in the hypothesis testing case, so we are not going to bother that much about the functional form. We are going to actually make use of this Engel curve to prove or disprove some hypothesis that we can check from the real life data.

So, a good Engel curve reflects its income elasticity and it indicates whether the good is an inferior, normal or luxury good. So, if you remember that I have spoken about in the previous lecture about the classification of goods and services in these 3 categories by looking at the income elasticity measures. So, the income elasticity can be obtained from the Engel curve and if the income elasticity is greater than 1 for a particular good or service, then it is called luxury.

If the income elasticity takes value less than 0, then we say that that particular good or service is an inferior commodity and if the income elasticity for demand for a particular good or service is taking fractional value, then actually we can say that it is a normal good. So, the border line between the luxury good and the normal good is basically the value 1. So, here we are going to take a case sometimes we see that poor household they do not spend that much of money on nutritious fruits and whereas you see the consumption of fruits is more visible and common in the upper middle class and the rich class of the society.

So, we may be interested to answer a particular socioeconomic question; so are fruits luxury items for some community? So, we can get some survey data, we can get data on the household consumption on different food items and we can get the data on income, if the households are not ready to release their income figures then you can ask their total monthly expenditure on all sort of consumption items and that will work as a proxy for income.

And then if you have got that kind of dataset, along with these kind of data you can also ask demographic data like size of the household, whether the household has children or not, how many children are there, et cetera. And then you have a very nice dataset with some explanatory variables and now you can start the journey of estimation of an Engel Curve and addressing this question of whether fruit is a luxury community or not.

So, here let us have a model. So, in my regression model I have log of the expenditure or fruit items, that is market dependant variable. Then I have log of income as one of the explanatory variables and then I have two other explanatory variables, one is household size and that is basically the number of adults in the households and then I have another explanatory variable child and that is the number of children in the households.

And suppose I got some dataset where I have 40 observations, these are all hypothetical arbitrary cases, so we do not have to worry why 40, it can be even 400. Anyway, so we have estimated the equation from the data and I am showing you the results in the table. So you see the coefficients are reported for the intercept term and the log of income variable and the household and the child variable as well. Standard errors are reported and then if you divide coefficient by the standard error, you are going to get the t statistic value. So, the t statistic values are also reported in the fourth column.

Now you have to conduct the t test and we are interested to know whether the fruits consumption is a luxury good or not. So, basically here we have to focus on the coefficient of the log income variable. And log income variable's coefficient, then we are going to test for a particular value of that coefficient. So, before we start the hypothesis in this context, let us look at the interpretation of these variables.

So, you see relationship between expenditure on fruits and the income is basically straight line in the log log plane, because we have taken log for both the variable. So, the regression coefficient associated with the log income variable is actually showing the elasticity value. And the value of that regression coefficient actually tells me that if there is 1 percent change in my income or household income, then by what percentage expenditure on fruits are going to change.

So, basically if you see that the regression coefficient takes a value 1, then actually it says that that is proportional increase. So if there is 1 percent change in income, then the household is going to spend exactly 1 percent extra on fruits and so on so forth. So, basically if fruits are indeed luxury items, then one could expect that value of the regression coefficient to be greater than 1. So, we are going to set the value for the null hypothesis that the coefficient will take value 1 and we can go for a right tail test, but here we are going for a two tail test.

So, here we write that null hypothesis is beta 1 is equal to 1 and null hypothesis, there is a typo here so it is alternative hypothesis. So, H1 will be the alternative hypothesis says that beta 1 is not equal to 1. So, we know the next step, we have to calculate the test statistic. And how do I get the test statistic?

So, the formula if you remember from the previous slide, you have to take the value of the observed coefficient which is 2.0045 here, you have to deduct value 1 which is the hypothesized value for the unknown population parameter beta 1 and then you have to divide that difference by the standard error for that log of income coefficient which is 0.5123 here.

So, if you do all this, then you get the calculated value of test statistic as 1.9607 and let us assume alpha to be 5 percent level and then as I said that we have 40 observations hypothetically, so here the degrees of freedom will be 36 and then you have to consult the t table to find the critical value. And the critical value from the t table is 2.0315. So, you see the critical value is higher than the test statistic value and hence you cannot reject the null hypothesis. So, as you cannot reject null hypothesis, you can say that well, the fruits are actually the luxury item.

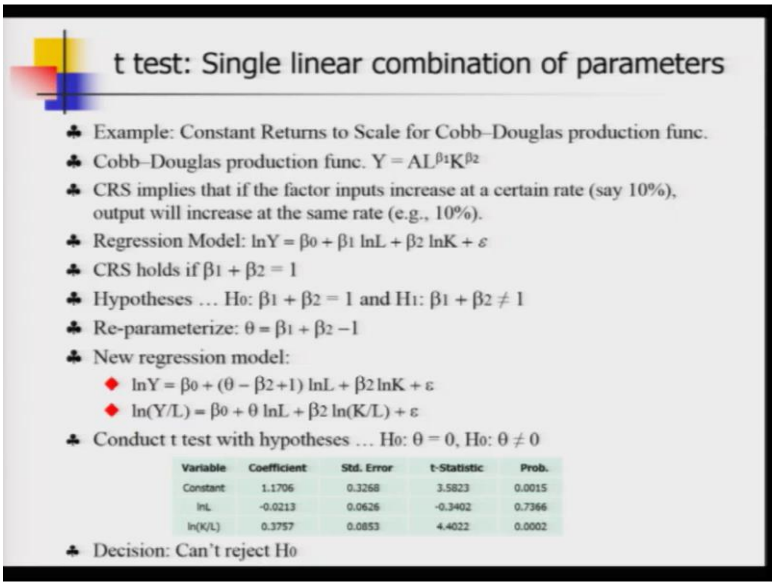
Now, we are going to discuss another case of t test and so far we have delt with one since parameter in the linear regression equation and we were interested in testing a particular value for that unknown population parameter. But what if we are interested in testing some relation between two or more unknown population parameters, can we do that via t test? So, here I am going to show you that yes, you can handle two regression coefficients in one linear combination setup and we are going to explain this by an example.

And for example again I am going back to the previous lecture. So, if you remember we have discussed the case of production function, and we discussed the case of a short term production function where we have one variable input and there could be many fixed inputs which are not varying. And now here in this case we are talking about a long run production function where 2 or more variable inputs are there, there is no fixed input and we have to assume specific form, a mathematical function to represent my production function so that for different levels of input I get the maximum value from that function, so that would give me the maximum amount of output which is possible to produce from that input bundle.

By input bundle I mean specific values of inputs that are there in the technology set or in the production function. So, here I am going to talk about the simplest possible two variable input production function case and I am going to assume the most popular production function form in empirical microeconomic research or in developmental research and that is known as the Cobb Douglas production function.

So, associated with this Cobb Douglas production function, and why Cobb Douglas? With all sort of production functions there is a concept called constant return to scale and we are going to test whether we observe constant return to scale from the sample data or not. So, in the next slide I am going to show you what is CRS and how we can test for CRS in the context of Cobb Douglas production function.

(Refer Slide Time: 34:57)



t test: Single linear combination of parameters

- Example: Constant Returns to Scale for Cobb–Douglas production func.
- Cobb–Douglas production func. $Y = AL^{\beta_1}K^{\beta_2}$
- CRS implies that if the factor inputs increase at a certain rate (say 10%), output will increase at the same rate (e.g., 10%).
- Regression Model: $\ln Y = \beta_0 + \beta_1 \ln L + \beta_2 \ln K + \varepsilon$
- CRS holds if $\beta_1 + \beta_2 = 1$
- Hypotheses … H_0: $\beta_1 + \beta_2 = 1$ and H_1: $\beta_1 + \beta_2 \neq 1$
- Re-parameterize: $\theta = \beta_1 + \beta_2 - 1$
- New regression model:
  - $\ln Y = \beta_0 + (\theta - \beta_2 + 1) \ln L + \beta_2 \ln K + \varepsilon$
  - $\ln(Y/L) = \beta_0 + \theta \ln L + \beta_2 \ln(K/L) + \varepsilon$
- Conduct t test with hypotheses … H_0: $\theta = 0$, H_0: $\theta \neq 0$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	1.1706	0.3268	3.5823	0.0015
lnL	-0.0213	0.0626	-0.3402	0.7366
ln(K/L)	0.3757	0.0853	4.4022	0.0002

- Decision: Can't reject H_0

So, the Cobb Douglas production function has this specific form that says Y equals to A times L to be power beta 1 times K to the power beta 2. And L is basically labour input, K is my capital input and beta 1 and beta 2 are the output elasticities with respect to labour input and capital input. And this parameter A, that is basically the technology parameter, so it says that the state of technology is fixed, you cannot change, that A is a parameter value it is given to you.

Now, what does constant return to scale implies in general? So, when you have a production function and you have variable inputs, so let me tell here that this constant return to scale is basically a long run concept where you know all factor inputs can be changed, there is no fixed input. So, in long run, you can change factor inputs proportionately. So, constant return to scale implies that if my factor inputs in production function are increased at a certain rate, then output will also increase at the same rate.

And if the output is increasing at a higher rate, then at the rate you are increasing your inputs then we call that there is increasing return to scale. And if you see that output is increasing at a lesser proportion, so at a lesser rate, then you can say that you are observing decreasing return to scale for your production function, based on the sample evidence that you have. So, now we have to test whether we can observe constant return to scale for Cobb Douglas production function fitted to a sample.

Now for that we have to first write down the linear regression model. So, if you note the mathematical form of Cobb Douglas production function, it is a non-linear function. So, we have to make this non-linear function a linear one. So, for that we have to take log both sides, and if we do so then we get a linear in parameter and linear in variable equation in terms of log Y log L and log K. Then you add the stochastic noise term epsilon and then you have a perfect linear regression model.

Now, in this linear regression model what situation actually indicates that the production function is displaying constant return to scale? So, if we look at the microeconomic theory, it tells us that if we add these two elasticity measures, output elasticity of labour and output elasticity of capital, if they sum to 1, then actually we have the case of constant return to scale in a two input Cobb Douglas production function. So we are going to test whether the sum of the exponents of the Cobb Douglas production function actually is 1 or not.

So, we start with hypothesis framing. So, our null hypothesis says that beta 1 plus beta 2 is equal to 1 and alternative hypothesis says that beta 1 plus beta 2 is not equal to 1. And then how to actually go about a t test in this case? There could be 2 different approachs and I am going to only talk about one particular approach and that is called the approach of re-parameterization.

So, here if you see that your hypothesis involves two parameters, beta 1 and beta 2, so you have to now do some re-parameterization of your original model or the mother model such that in the null hypothesis you get only one particular parameter and then you can conduct your t test. So, that is what we are going to do now.

So, next we re-parameterize and now let me introduce a new parameter theta, that is equal to beta 1 plus beta 2 minus 1. And if I now make use of this new parameter, then I can rewrite my regression model and that you see the first red diamond bullet in the slide, so I am writing log of Y equals to beta 1 plus theta minus beta 3 plus 1 times log L plus beta 3 log K.

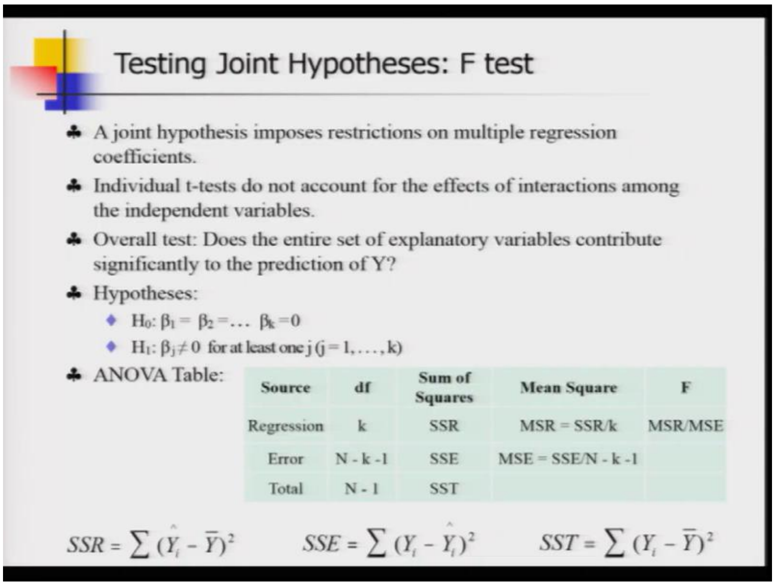
So, you see, I have actually got rid of the beta 2 parameter altogether from my linear regression equation and now I can rewrite it further. So, basically I can also write log of Y divide by L, so for that actually I have to divide this by log of L and then basically I have what, beta 1 plus theta times log of L plus beta 3 times log of K divided by L plus epsilon. So, that is basically my right hand side.

So, this is my new linear regression model. And note that I am going to run my OLS technique on this re-parameterized model. So, based in the data I am now going to get the estimates for theta and beta 1 and beta 3 but not for beta 2.

So, now you conduct your t test with hypothesis and it is a re-parameterized hypothesis of course. So, theta we are testing for theta's value of 0 against the alternative hypothesis of theta is not equal to 0. And here, I am showing you simple regression table. All numbers are fictitious, so basically you have the intercept or constant term and log of L and log of K by L's coefficients are also reported. Now you are going to make a decision, so you have to now look at the theta parameter and the corresponding t statistic value and the corresponding p value.

So, let us now look at the table, so here the theta parameters estimated value is minus 0.0213 and the t statistic value is very small, it is minus 0.3402 and you see the associated probability value is very high, 0.7366. So, now if you choose standard level of significance alpha equal to 0.05, so here you can see that the probability value if much, much higher than the generally accepted level of significance. So, you cannot reject the null hypothesis.

(Refer Slide Time: 42:08)



Testing Joint Hypotheses: F test

- A joint hypothesis imposes restrictions on multiple regression coefficients.
- Individual t-tests do not account for the effects of interactions among the independent variables.
- Overall test: Does the entire set of explanatory variables contribute significantly to the prediction of Y?
- Hypotheses:
  - H_0: $\beta_1 = \beta_2 = \ldots \beta_k = 0$
  - H_1: $\beta_j \neq 0$ for at least one j ($j = 1, \ldots, k$)
- ANOVA Table:

Source	df	Sum of Squares	Mean Square	F
Regression	k	SSR	MSR = SSR/k	MSR/MSE
Error	N - k -1	SSE	MSE = SSE/N - k -1	
Total	N - 1	SST		

$SSR = \sum (\hat{Y}_i - \bar{Y})^2$ $SSE = \sum (Y_i - \hat{Y}_i)^2$ $SST = \sum (Y_i - \bar{Y})^2$

So, what does the joint hypothesis tells us? So, it imposes restrictions on multiple regression coefficient. And individual t tests are not going to be of any help here because they do not account for the effects of the interactions among the independent variables. So, here is an interesting point. So, I have a multiple regression equation model now where I have K number of explanatory variables. And in reality there could be some linear correlation between these explanatory variables. So, there could be positive or negative covariance value between these variables.

So, what I actually want to say that the beta coefficients that I am computing via OLS method may not be independent of each other. There could be some connection, there could be some covariance or correlation between the estimated beta coefficients. Because do not forget that beta coefficients themselves are random variables because they are linear combinations of normally distributed error variables.

So, we cannot just conduct individual t test. Suppose we have a regression model where we have 5 explanatory variables, can I conduct 5 separate t tests for different slope coefficients and then say I got 3 significant, 2 insignificant and hence, I say that overall some variables are helping me to model the variation in Y. No, you should not actually do that, actually you have to go for a overall or joint test and that is what we are going to see how f test is going to help us to conduct the same.

So, here what do we mean by overall test? So, we pose this question, does the entire set of regressors contribute significantly to the variation or prediction of y. So, when we are going to address these question or concern, we are going to make use of joint hypothesis testing or that is also called overall testing. So, we start with hypothesis, our null hypothesis says that jointly beta 1 beta 2 dot dot dot beta k are all equal to 0 and alternative hypothesis says that beta j is not equal to 0 for at least one j, where j is anything from 1 to k.

So, what does my alternative hypothesis is saying? So, alternative hypothesis is saying that out of 5, 6, 7 whatever number of explanatory variables that you have in your model, at least one, I don’t know which one, but I can say that at least one is helping me to explain some bit of variation in y and hence, the associated regression coefficient is not equal to 0. So, there is some linear relationship between that particular explanatory variable and the dependant variable.

Now we are going to make use of our good old friend of ANOVA table and anova table I already have shown to you before, so it is nothing new so I am not going to describe, I am just only going to say that there is one concept called ANOVA identity which helps us to partition the variation in Y in two components, component number 1 is basically the variation which is explained by my regression model and then there is another component which remains unexplained. So, the part of the variation in Y which is not explained by my regression model.

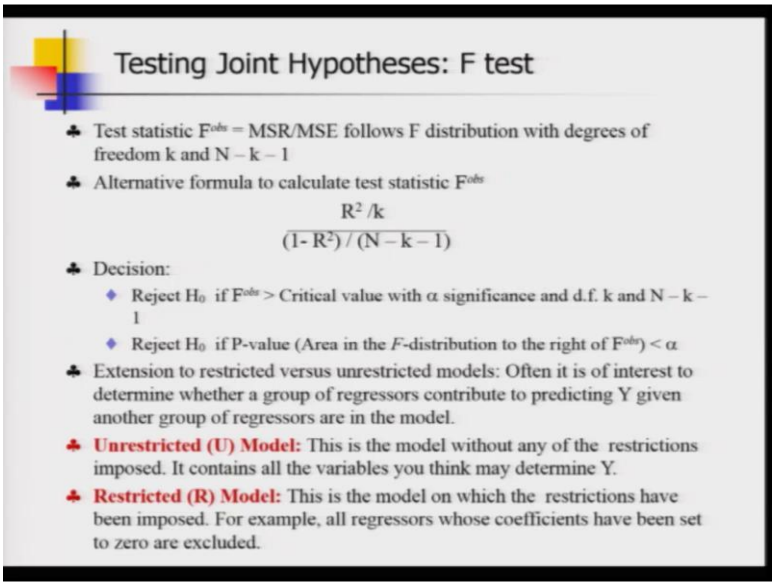
So, here in the ANOVA table you see there are 3 rows under the column source, so the total is basically talking about the total variation in my dependant variable and there are two components, one is basically regression and the other one is error. So, regression component is basically saying that how my fitted regression line is helping me to explain some bit of variation in my dependant variable y.

So, for that the formula I am showing here at the bottom of the slide, if you have forgotten it, and this will probably remind you about the calculation for SSR that we know we have done previously. And here we get basically the difference of the fitted value and the mean of the fitted value and it can be shown theoretically that the mean of the fitted value Y hat i is Y bar itself. So, basically you take the difference of fitted values from it means and then from its mean and then you square the difference and then sum and that is the way you get the SSR.

And SSE is also not unknown to you. This is the entity that we actually minimise in order to get my OLS coefficient estimates. And I have explained previously that they are at degrees of freedom associated with this sources and of course, if you divide the sum of squares that comes from each of these components regression and error, you can divide them by the corresponding degrees of freedom and that is the way you get the mean square and that is basically MSR and MSE for the regression component and the error component respectively.

So, here in the column I am also showing he MSR and MSE must be calculated and then finally once the mean square numbers are with you, you take the ratio of MRR and MSE and that is your f statistic and then basically from that f statistic you can conduct f test.

(Refer Slide Time: 48:26)



Testing Joint Hypotheses: F test

- Test statistic F^{obs} = MSR/MSE follows F distribution with degrees of freedom k and N – k – 1
- Alternative formula to calculate test statistic F^{obs}

$$\frac{R^2/k}{(1-R^2)/(N-k-1)}$$

- Decision:
  - Reject H_0 if F^{obs} > Critical value with α significance and d.f. k and N – k – 1
  - Reject H_0 if P-value (Area in the *F*-distribution to the right of F^{obs}) < α
- Extension to restricted versus unrestricted models: Often it is of interest to determine whether a group of regressors contribute to predicting Y given another group of regressors are in the model.
- **Unrestricted (U) Model:** This is the model without any of the restrictions imposed. It contains all the variables you think may determine Y.
- **Restricted (R) Model:** This is the model on which the restrictions have been imposed. For example, all regressors whose coefficients have been set to zero are excluded.

So, this f statistic which is denoted by f observed here follows an f distribution with two degrees of freedom k and n minus k minus 1. And note that although all statistical software reports ANOVA table, when they are reporting regression coefficients, but if we know you are not given ANOVA table, can you still conduct the f test by looking at something else? Yes, you can.

Actually there is a very nice relationship between SSE and the R squares. And by making use of the relationship between ANOVA identity and the R square formula we can alternatively propose formula to calculate the test statistic value f observed, and that is given by R square divided by k and that the entire thing is divided by 1 minus R square divided by N minus K minus 1. It is little bit complicated, but you should make use of the ANOAVA table as the similar alternative.

Now, how do you, to accept or reject your null hypothesis? So it simple, you reject your null hypothesis if the observed test statistic value is higher than the critical value with alpha significance level and degrees of freedom K and N minus K minus 1. So, if you want to make use of the p value in order to take your decision, then you have these following decision rule. You reject null hypothesis if the p value, which is basically the area in the f distribution to the right tail of observed value of f statistic. If it is less than alpha, the chosen level of significance, then you reject your null hypothesis.

Now, this f test can be used to deal with many other types of hypothesis testing problems and one of the most important hypothesis testing problem is the context of restricted versus unrestricted models. But today, I do not have enough time to go through it. So, in the next lecture I am going to start the discussion with the restricted versus unrestricted model case and how f test can dissolve this issue. So, come back for the next lecture. Thank you, see you then, bye.