Exploring Survey Data on Health Care

Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee

Lecture - 33
Fixed Effect Model in Healthcare

Welcome friends, once again to my NPTEL MOOC module on Exploring Healthcare Data, Healthcare Survey Data. As part of the Minister of Educations Initiative of the NPTEL program where so many unique courses are being floated, I have tried my best to contribute to this gap in different courses given the challenges of physical education.

This time we have considered the model on healthcare, the course outline is on healthcare and understanding healthcare data. These are called healthcare data analytics as well. Here we are dealing with 7th-week lectures. In the 7th week, we have already taken two lectures where we clarified what are called panel data, how to understand or read panel data, and how to know to balance and unbalance panel data.

In the previous lecture, we talked about the common constant model or pool panel model. So, even if it has panel content that is regarded as a cross-section type. In this lecture, we are adding feathers to the existing information on the panel that is a dimension called the effect, fixed effect model.

I am just clarifying the overview of it fixed. The word fixed effect means in the multidimensional set of with time and cross-sectional variation as part of the panel structure of the data. Some variables by structure are defined to be changing with a fixed percentage, with a fixed component, fixed coefficient.

The fixed coefficient may not be captured in the slope coefficient. Slope the change is not exactly captured, but the constant term could be actually identified as the change. Like when you say your consumption pattern every year, it might be hardly changing with different variables.

if your location is going to get changed, I am just giving the various type of examples. Location is changing, your attire is changing, your you are putting on a new dress, for example, your consumption is not going to change much. Even if your income is also changing the extent of consumption is changed by a certain degree due to a certain dimension of the product,

Some of the yes consumption is a function of income and income change has reflection on consumption for sure. But there is some parallel example in this case though you are a bit confused with this. Some parallel examples like a person's, person height over the time due to a factor of various, as a function of various factors, or consumption as a person's caste for example caste over a period of time.

Caste, religion, etcetera within a family is not going to be changed over time or over a period of time, it is not getting changed much. But when you just try to capture the time effect probably you are not going to identify the exact aspect of the change. But that does not mean it is not changing with a structure, the alpha that is the constant term is somewhere reflecting the standard of living of the person,

No, it is not; though exactly capturing the variation in terms of time, but there might have been a paradigm shift that is captured in terms of the constant term that is an alpha. So, that is where the panel data stands. And the panel data clarify through the fixed effect on the constant changes, on some unit of v top in a change in the constant term, not by the slope coefficient,

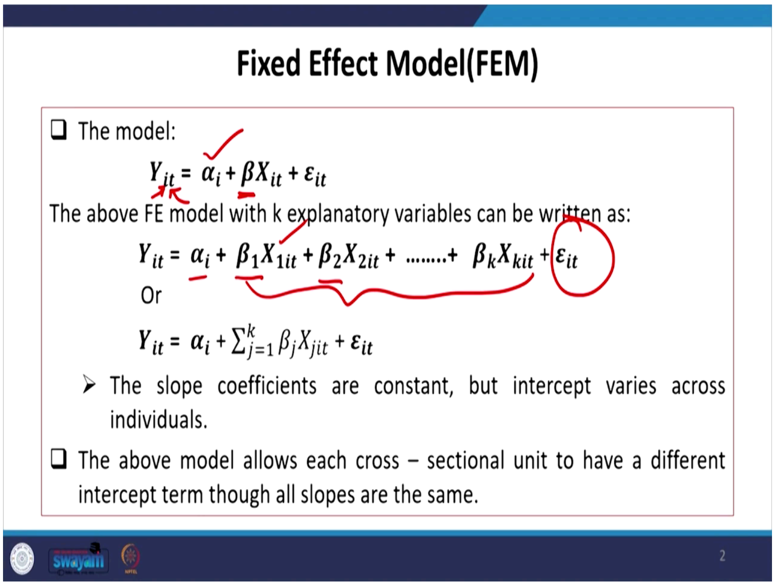
That is why we are dealing with this particular lecture and emphasizing on the fixed effect of the panel content. And this is going to be very interesting in your everyday work, especially those who are targeting for the topmost journal, topmost paper, panel paper, and panel data used to be very important to ice icing your model, to give a better direction in your model, to have a better interpretation. So, far as policy recommendation is concerned. That is one direction I said.

Then the second aspect is whoever is doing some kind of an analysis using the large scale data; large scales cross-sectional data. But it has a time component as well. It is a kind of microdata, but the time horizon is very less. It is a kind of sort panel not a long panel where your cross-sectional dimensions are there, but not with by the time.

So, cross-sectional dimensions are obviously, capture certain variations in turn dynamics, So, the within effect is not so significant, but between effect might be there. So, that between effect and within effect clarification I give it in the previous lecture. So, that aspect you can easily capture through the fixed effect model,

So, the fixed effect model is of course, useful in research and I suggest strongly that you please follow the line, between the space, between the subscript and superscript and understand it very clearly and you will enjoy it like anything, So, let us go ahead and clarify the equation once again. Starting with the very basic fundamental equation of the panel data. The dependent variable is with the subscript i and i that you can easily see.

(Refer Slide Time: 06:43)



Fixed Effect Model(FEM)

❑ The model:

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$$

The above FE model with k explanatory variables can be written as:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + ...+ \beta_k X_{kit} + \varepsilon_{it}$$

Or

$$Y_{it} = \alpha_i + \sum_{j=1}^{k} \beta_j X_{jit} + \varepsilon_{it}$$

➢ The slope coefficients are constant, but intercept varies across individuals.

❑ The above model allows each cross – sectional unit to have a different intercept term though all slopes are the same.

I am clarifying bit by bit and giving you the right direction. You might not have received this particular information from my previous year's model and previous model on handling large-scale data. I did not explain much. But now I feel that based on the feedback of the students, I feel that re-emphasis should have been there on each of the coefficients.

Starting with fixed content, then with the slope coefficients etcetera. Now, you can see that this beta term is a is not going to change much, it is constant. So, that means, our by slope by slope dimension it is not identify the changes, but overall the model is going to capture, the certain dimensions in terms of cross-sections are captured in the constant term,

Therefore, it has certain assumptions, So, that the same equation is further elaborated as alpha. So, beta is beta 1 plus beta 2, this is not beta i, each of the beta coefficients is now estimated with its control variables. And these are all our control variables, and this is our error term. An error term is, of course, attached with each of the dimensions, cross-section dimensions as well as time dimensions.

So, I just said that the slope coefficients are going to be constant, but the intercept varies across individuals not across time. The time component is not bearing the model as such significantly rather it is differentiating the model with its alpha term.

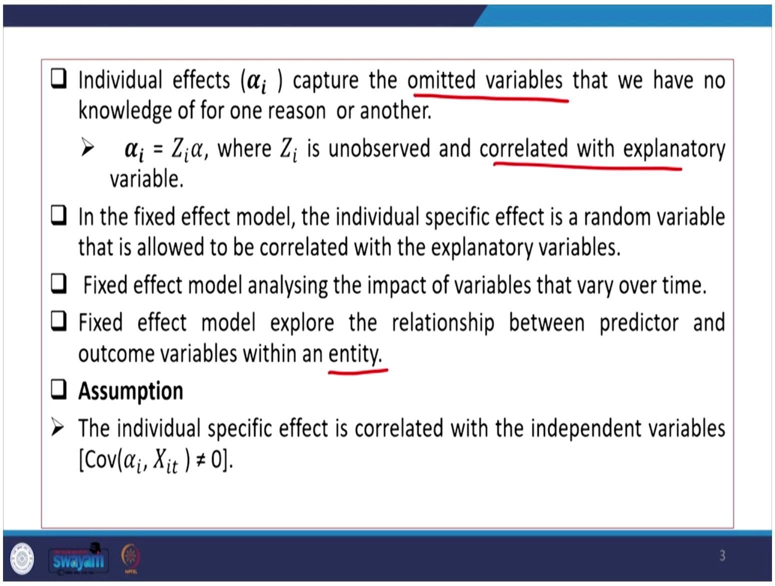
But in the common constant model alpha is actually constant. in the cross-section model or even the ordinary release square model, you could have noticed that alpha is not alpha i, Here alpha i is observed because of the panel content. on the aggregate where you have data, all are cross-sections, there is no question of the same person is indicating or giving certain views over time.

But the same persons when repeated even if you are considering minus the t term, but the same persons have certain cross-sections as well, cross-section variations within the time period as well. So, that is why the alpha term is actually capturing the dynamics of the cross-sections,

So, in the early earlier cases, the alpha term does not consider the dimensionalities of its cross-sectional changes. The above model allows each cross-sectional unit to have a different intercept term through all slopes through all slopes are the same. That is what is called individual effect,

Alpha is in fact, called the individual effect. Each individual's dimensions are actually captured. And this captures the omitted variables. if there are any sort of variables omitted by random procedure or by non-random procedure, in both the case omitted variables have a possibility of dropping or the possibility of deletion that is actually captured through the individual effects. That is why it is called alpha i.

(Refer Slide Time: 10:08)



❑ Individual effects ($\boldsymbol{\alpha_i}$) capture the omitted variables that we have no knowledge of for one reason or another.

➢ $\boldsymbol{\alpha_i}$ = $Z_i\alpha$, where Z_i is unobserved and correlated with explanatory variable.

❑ In the fixed effect model, the individual specific effect is a random variable that is allowed to be correlated with the explanatory variables.

❑ Fixed effect model analysing the impact of variables that vary over time.

❑ Fixed effect model explore the relationship between predictor and outcome variables within an entity.

❑ **Assumption**

➢ The individual specific effect is correlated with the independent variables [Cov(α_i, X_{it}) ≠ 0].

So, the most important aspect for all of you to note is that omitted variable bias is actually captured in the case of the fixed-effect model. And that we have no knowledge of for one reason or the another about capturing this individual effect and that is captured. Then what is this alpha i?

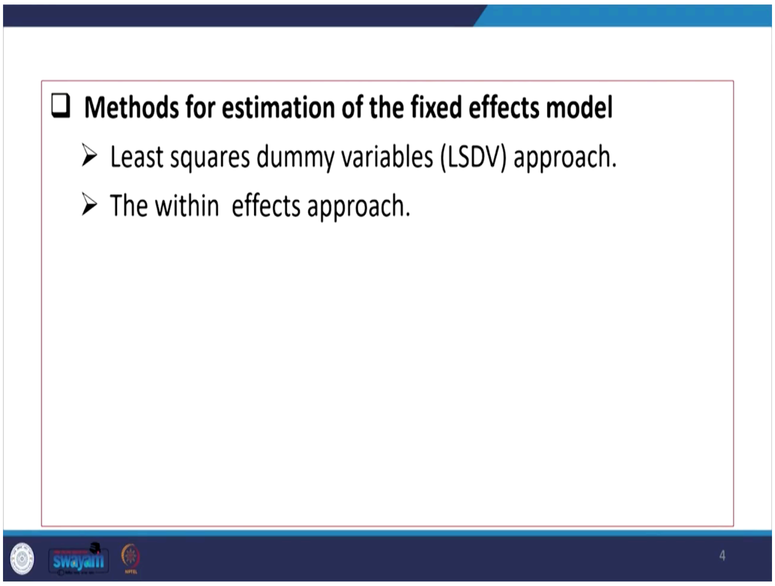
Alpha i, in fact, is a proportion of the alpha that is through Z i, Z i is actually unobserved. And since it is unobserved consider to be correlated with the explanatory variables, So, that is why there are some problems with the OLS model. So, the normal OLS model is not going to be applied since the Z i is correlated with the explanatory term.

In the fixed-effect model, the individual effect or specific effect is a random variable that is allowed to be correlated with the explanatory variables. Fixed effect model analyzing the impact of variables that vary over time. The fixed-effect model explores the relationship between predictor and outcome variables within an entity.

So, each individual and it is effects are actually captured that is why we say within an entity. And that actually occurs due to the time changes, time dimension, though time-wise there are also changes, in the individual case there are expected to be some variations.

So, there, so the assumption in the specific individual effect is correlated with an independent variable which is why the covariance of alpha i and x is nonzero, and this is a serious error in OLS. So, OLS is not applied.

(Refer Slide Time: 11:55)

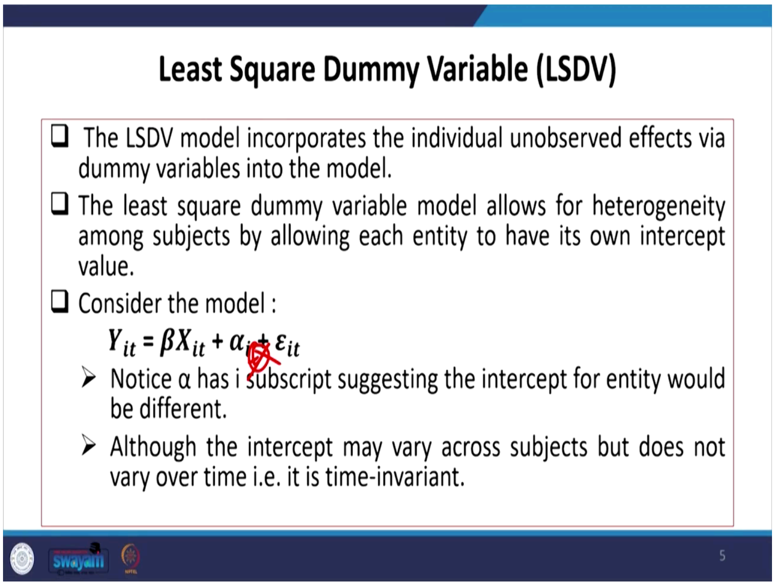


- ❑ **Methods for estimation of the fixed effects model**
  - ➢ Least squares dummy variables (LSDV) approach.
  - ➢ The within effects approach.

So, the methods for estimation of the fixed effect model are made through two approaches one is through LSDV model that is famously called LSDV or called Least Square Dummy Variables approach, and the second one is called within effects approach, Now, we are going to experiment with the LSDV approach to take certain variables to be with a dummy content with the dummy variables model.

And we will clarify how those problems can be avoided and then we can apply for an OLS model, The LSDV model incorporates the individual unobserved effects via dummy variables into the model.

(Refer Slide Time: 12:42)



Least Square Dummy Variable (LSDV)

- ❑ The LSDV model incorporates the individual unobserved effects via dummy variables into the model.
- ❑ The least square dummy variable model allows for heterogeneity among subjects by allowing each entity to have its own intercept value.
- ❑ Consider the model :

  $Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$

  - ➢ Notice α has i subscript suggesting the intercept for entity would be different.
  - ➢ Although the intercept may vary across subjects but does not vary over time i.e. it is time-invariant.

The least-square dummy variable model allows for heterogeneity among subjects by allowing each entity to have its own intercept value, So, through the dummy variables each entity has its own intercept value. Let us consider the model once again. This is equal to beta times X it. So, beta is constant here, but now the other two terms are alpha i and E it is actually having certain variability, but t is not there. So, t if I say it, it is actually not there,

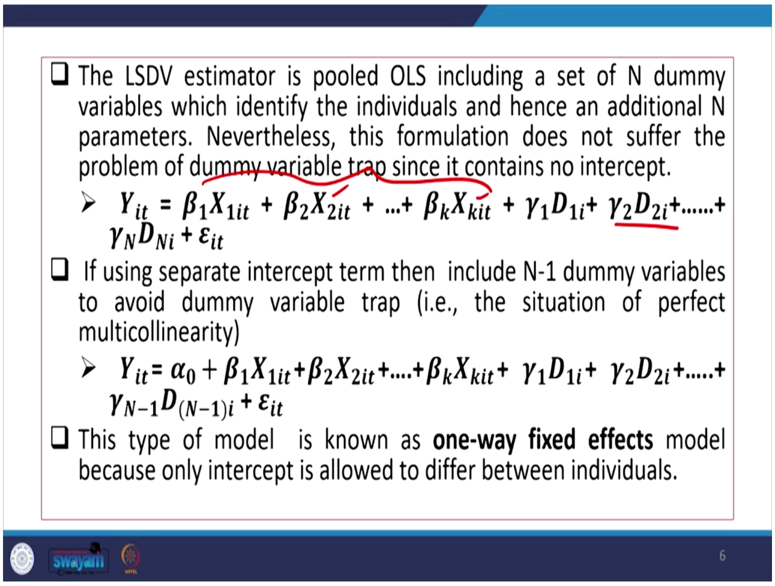
So, t component is not there, but variability is observed through the individual variations and due to what is called if someone observes effects are not captured, so that can be captured through the alpha i. Although the intercept may vary across objects, but does not vary over time. So, that is the reason why it is called time-invariant model. This is also called time invariant. So, there is not term attached with the alpha i, That is why this is called time invariant.

So, the LSDV estimator is actually called a pooled OLS including a set of N dummy variables. This identifies the individuals and enhance an additional N parameter, since we are adding dummies N dummies, So, N parameters are again generated.

Nonetheless, this formulation does not suffer the problem of dummy variable trap since it contains no intercept. We have already read earlier about what is called dummy variable trap. If you take parameters as compared to the number of categories, if they are the same then it is going to have certain problems in the model, that kind of problem of estimation is called a dummy variable trap.

So, the dummy variable trap is avoided depending upon the alpha content we are taking. Like, in this in the first equation Y it is equal to the first beta 1 till X i term.

(Refer Slide Time: 15:02)



❑ The LSDV estimator is pooled OLS including a set of N dummy variables which identify the individuals and hence an additional N parameters. Nevertheless, this formulation does not suffer the problem of dummy variable trap since it contains no intercept.

➢ $Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \ldots\ldots + \gamma_N D_{Ni} + \varepsilon_{it}$

❑ If using separate intercept term then include N-1 dummy variables to avoid dummy variable trap (i.e., the situation of perfect multicollinearity)

➢ $Y_{it} = \alpha_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \ldots\ldots + \gamma_{N-1} D_{(N-1)i} + \varepsilon_{it}$

❑ This type of model is known as **one-way fixed effects** model because only intercept is allowed to differ between individuals.

Now, the alpha unit is expressed into N into a N number of dummies since alpha is not there. So, that one is gamma times 1 D 1 i. First, i is captured, then the second like one X i X 2 i, X 3 i etcetera. Here we are defining each dummy till N dummies to replace that alpha i just to identify it its coefficient so that we can estimate the right coefficient in this model.

If using separate intercept term then in intercept is there and so N minus 1 dummy has to be considered to avoid the dummy variable trap issues, Then, there will be any if you do not consider then dummy variable trap is actually explained through the multi-collinearity problem and multi-collinearity is going to give you 1 as the value, because that there will be a perfect multi-collinearity.

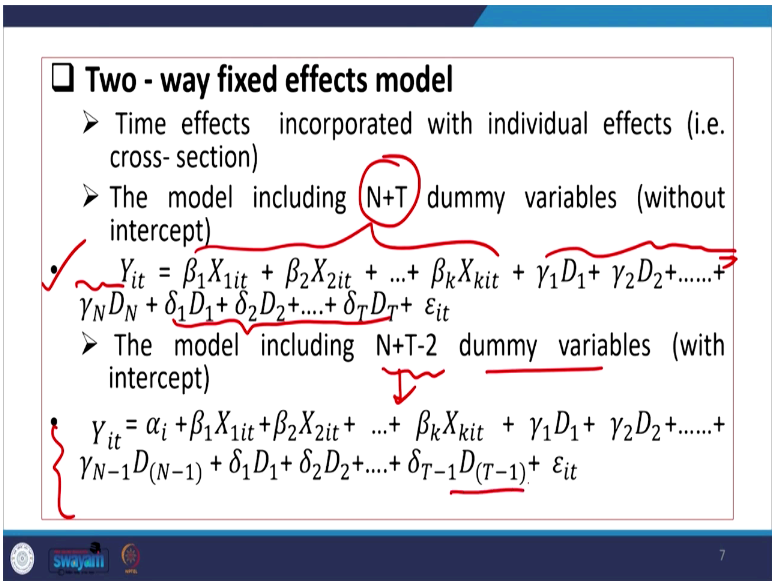
If you have some difficulties here reading the dummy variable trap, I suggest you to read the simple text bo called by Damodar Gujarati. It will give you the right directions as well to explain to you a dummy variable trap correctly.

So, in this one when we are considering the alpha as the coefficient alpha term, then basically the that since this is there to avoid dummy variable trap we have considered N minus 1 dummy, So, N minus 1 dummies, then each of the estimations is going to give you the correct prediction of the model.

So, this type of model is also known as a one-way fixed-effect model, because only the intercept is allowed to differ between individuals. So, since intercept term is now actually allowed to be estimated through different intercepts that is why I called the one-way fixed effect.

So, obviously, there will be a fixed two-way, fixed effect model, in that case, time effects can be also incorporated with the individual effect, So, like the model includes N plus T dummy variables as well if time effects are also captured.

(Refer Slide Time: 17:27)



❑ **Two - way fixed effects model**

- Time effects incorporated with individual effects (i.e. cross- section)
- The model including N+T dummy variables (without intercept)

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \gamma_1 D_1 + \gamma_2 D_2 + \ldots\ldots + \gamma_N D_N + \delta_1 D_1 + \delta_2 D_2 + \ldots. + \delta_T D_T + \varepsilon_{it}$$

- The model including N+T-2 dummy variables (with intercept)

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \gamma_1 D_1 + \gamma_2 D_2 + \ldots\ldots + \gamma_{N-1} D_{(N-1)} + \delta_1 D_1 + \delta_2 D_2 + \ldots. + \delta_{T-1} D_{(T-1)} + \varepsilon_{it}$$

7

So, again without the intercept then there will be N plus T minus 2 dummy because now we have time dimensions as well. So, first of all, when time dimensions are there, this is our basic model. Now, this is without the time, so till N now you can see this is continuing to this till N without the alpha term um.

Now, we have the time dimensions, So, since so dummies are considered to be defined N minus 1 and N minus, in this case, N minus 1 and T minus 1 and so this is basically equal to N plus T minus 2 dummy variables, alright, with the intercept term if you are including the intercept term. If we do not have an intercept term then it will be simply N plus T dummy because the intercept term is not there.

This equation basically is without the intercept term. So, in this equation the dummy variables are N plus T minus N minus 1 and T minus 1 that is N plus T time minus 2 dummies are there. And accordingly, we can have captured each of the differences. So, this is called a two-way fixed-effect model.

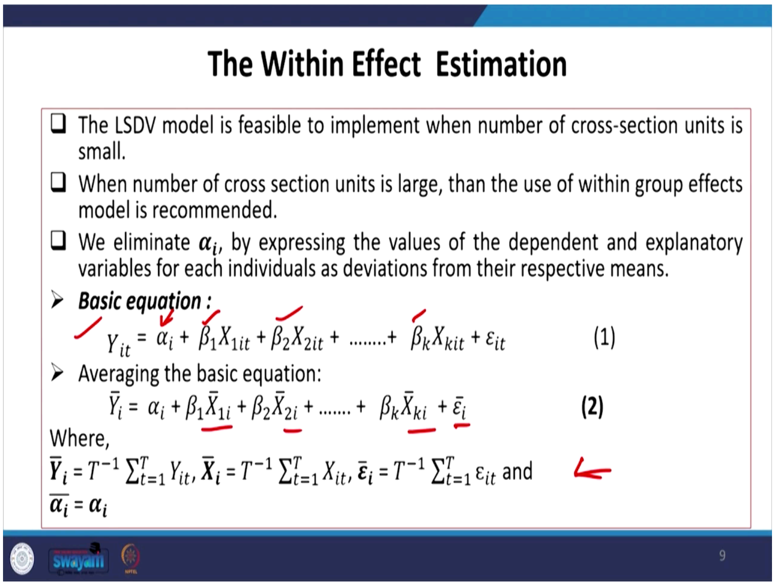
There are some advantages and disadvantages. The advantages are it has simplified the model. The disadvantage is that there are too many dummy variables if you have a pool of cross-sections. If so many cross-sections are there then you might have so many dummy variables that lead to a loss of degrees of freedom.

And this may not be able to identify the impact of time in the variant variable. Many dummy variables will lead to the problem of multi-collinearity. So, that is there.

Now, I am also explaining what is called the within-effect estimation. Within effect first, we have explained the concept called LSDV, what is called the least square dummy variable model. We will have practical sessions. Once I explain this within effect estimation I will go for the practical session or add certain results with the original data and we will clarify.

The LSDV model is feasible to implement when the number of cross-section units is small, When the number of cross-section units is large then it is better not to use the LSDV but rather to use a within-group a fixed model that is through the within effect model.

(Refer Slide Time: 20:19)



The Within Effect Estimation

- ❑ The LSDV model is feasible to implement when number of cross-section units is small.
- ❑ When number of cross section units is large, than the use of within group effects model is recommended.
- ❑ We eliminate $\boldsymbol{\alpha_i}$, by expressing the values of the dependent and explanatory variables for each individuals as deviations from their respective means.
- ➢ ***Basic equation :***

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots\ldots + \beta_k X_{kit} + \varepsilon_{it} \qquad (1)$$

- ➢ Averaging the basic equation:

$$\bar{Y}_i = \alpha_i + \beta_1 \bar{X}_{1i} + \beta_2 \bar{X}_{2i} + \ldots\ldots + \beta_k \bar{X}_{ki} + \bar{\varepsilon}_i \qquad \mathbf{(2)}$$

Where,

$\bar{\boldsymbol{Y}}_{\boldsymbol{i}} = T^{-1} \sum_{t=1}^{T} Y_{it}$, $\bar{\boldsymbol{X}}_{\boldsymbol{i}} = T^{-1} \sum_{t=1}^{T} X_{it}$, $\bar{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ and

$\overline{\boldsymbol{\alpha_i}} = \boldsymbol{\alpha_i}$

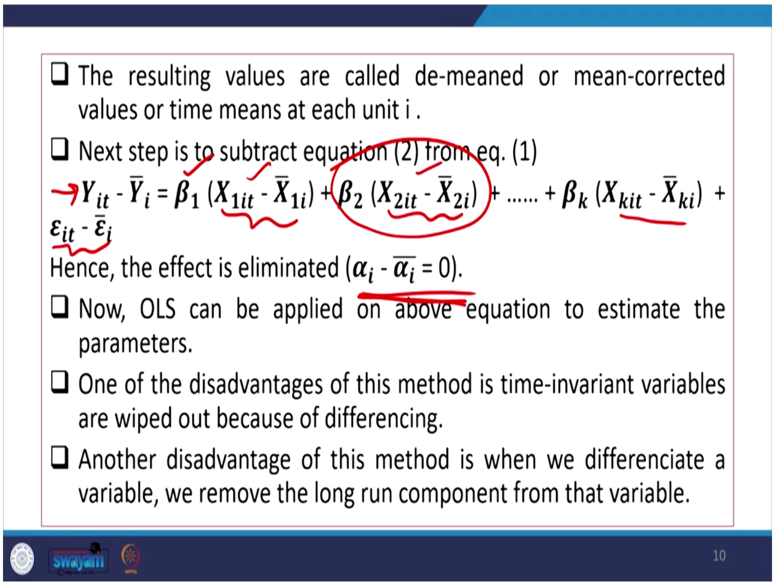
We eliminate alpha i by expressing the value of the dependent and explanatory variables for each of the individuals as deviations from their respective means, When we know that the individuals are not actually varying, some of the variables are actually not varying much or not at will varying, and some characteristics of panel data that is some of are not varying over time.

So, if you just deviate their mean value from the individual entries, then that will be defined, which will be considered to be having 0 variations. So, most of your variables are going to be having 0 values, and that way we can avoid many of the problems in the model.

So, the equations as per I just addressed are that this is our first panel data with slope coefficients are constant and then this is the constant term with certain variations, cross-sectional variations. Now, we are trying to take the averages, average overtime, and overtime of each of the explanatory variables. Now, what it is going to give us is that this is your average X bar, 1 i, then 2 i, then k i, alright. So, this is also average over time,

Now, we have taken the difference between the first equation and the second one. So, here it has the next equations to clarify what is this average. Average is basically over the time only T tends to 1 to T in each of the cases and you can find out the average accordingly.

(Refer Slide Time: 22:16)



□ The resulting values are called de-meaned or mean-corrected values or time means at each unit i .

□ Next step is to subtract equation (2) from eq. (1)

$Y_{it} - \bar{Y}_i = \beta_1 (X_{1it} - \bar{X}_{1i}) + \beta_2 (X_{2it} - \bar{X}_{2i}) + \ldots\ldots + \beta_k (X_{kit} - \bar{X}_{ki}) + \varepsilon_{it} - \bar{\varepsilon}_i$

Hence, the effect is eliminated ($\alpha_i - \overline{\alpha_i} = 0$).

□ Now, OLS can be applied on above equation to estimate the parameters.

□ One of the disadvantages of this method is time-invariant variables are wiped out because of differencing.

□ Another disadvantage of this method is when we differenciate a variable, we remove the long run component from that variable.

The resulting values are called demeaned values or time means at each unit i; each by each unit i we have actually derived the mean values. So, the next step is for us to subtract these two equations from 2 to 1, So, 2 is our or 1 to 2 from or from we are subtracting from the equation 1 to 2; 1 is here and then subtracting these two. What is going to give us is that like we have taken Y it minus Y it minus this Y i bar, so in each case, we have got the difference right.

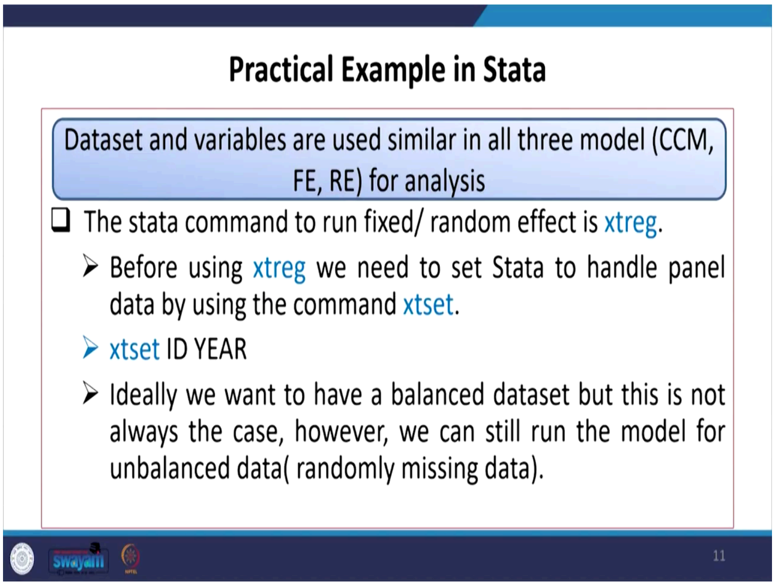
Now, what is interesting is that if there are no such changes in the constant term especially when we know that it is having a certain fixed effect, fixed within effect; within effect is fixed, in that case, alpha i and alpha bar is going to be 0. When this is going to be completely 0, then the rest of the equation is simply of our OLS type of model, .

It is simply having beta constant values or beta coefficient having no further changes within it or dynamics within it. So, we will simply estimate the beta values, So, it is not beta i, it is only beta as a beta 1, beta 2, and beta k. So, each of the parameters we can easily estimate since now alpha i is avoided. Now, we will, therefore, use the OLS in the equation to estimate the parameters. One of the disadvantages of the model is the time-invariant variable.

Some of the time-invariant variables are wiped out because of differentiating. Even X i we say X 1 till X i 1 i bar, X 2 i bar etcetera it may be the case that one of these coefficients might be wiped out because of its constant variation, all the variations are constant for the same categories or some coefficient variable. So, that value will be also difficult to estimate because that is time-invariant.

So, another disadvantage is that this method is creating some problems in terms of differentiating. When we differentiate a variable we remove the long-run component from that variable. So, in the long run the time component is actually completely also wiped out. So, the time dimensionality is not going to be interpreted at all in this model. So, now, we have the practical examples and we are trying to show you the results with the help of the data.

(Refer Slide Time: 25:17)



Practical Example in Stata

Dataset and variables are used similar in all three model (CCM, FE, RE) for analysis

- The stata command to run fixed/ random effect is xtreg.
  - Before using xtreg we need to set Stata to handle panel data by using the command xtset.
  - xtset ID YEAR
  - Ideally we want to have a balanced dataset but this is not always the case, however, we can still run the model for unbalanced data(randomly missing data).

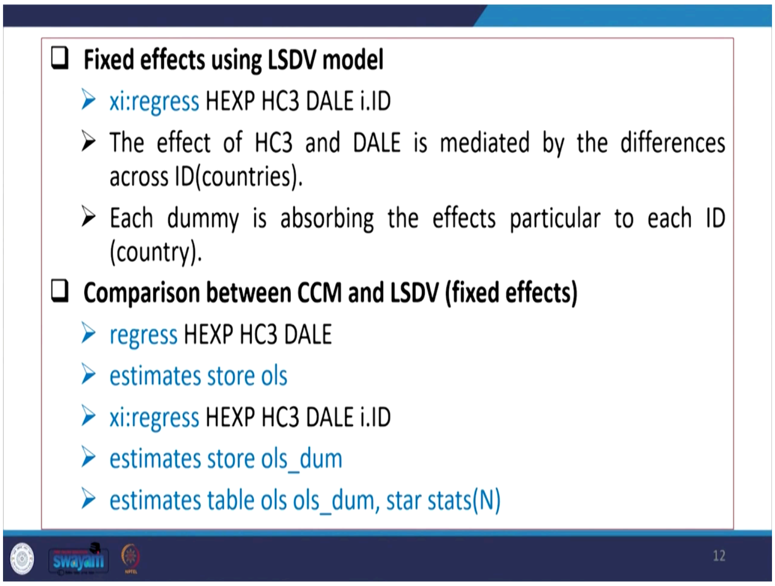
The data set and variables are used similarly in all 3 models that are CCM, fixed effect as well as a random effect. The random effect model we are going to go deal with in our next lecture. At this moment we are comparing CCM and FE. FE stands for fixed effect. Now, you might be confused that how to get this data, if I have data how to develop a panel. I will be using WHO OLS IHDS panel data.

We have already shown you how it is defined to be a panel and whether it is strongly a panel or a strongly balanced panel or not, Rest of the details you want to develop your panel data you can refer to my module that was earlier developed which is called handling large-scale data set with stata.

So, the stata command to run fixed effect and random effect is your xtreg, xt regression; xt is often used in the cache, it is used in cache or panel data estimation before using xtreg we need to set stata to handle the panel. So, we need to set the stata we need to set or set the stata with the command xtset to set the stata about to recognize the stata about your data whether it is a panel or not.

So, we want to also understand whether it is a balanced one or not a balanced one, We can still run the model for unbalanced panel data with certain missing observations as well.

(Refer Slide Time: 26:53)



❑ **Fixed effects using LSDV model**

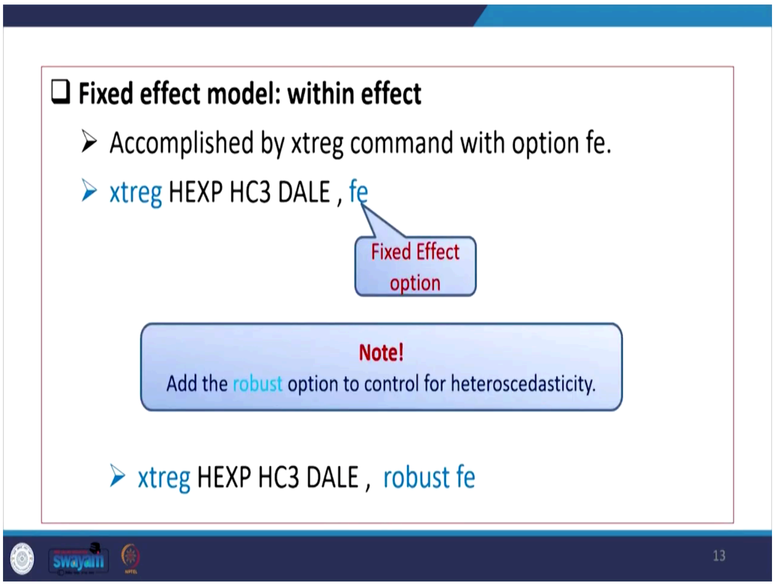
- xi:regress HEXP HC3 DALE i.ID
- The effect of HC3 and DALE is mediated by the differences across ID(countries).
- Each dummy is absorbing the effects particular to each ID (country).

❑ **Comparison between CCM and LSDV (fixed effects)**

- regress HEXP HC3 DALE
- estimates store ols
- xi:regress HEXP HC3 DALE i.ID
- estimates store ols_dum
- estimates table ols ols_dum, star stats(N)

Now, here are all our commands and I am going to show them one by one. We will also compare the fixed effect using the LSDV model, then a comparison between CCM and LSDV fixed effect model as well.

(Refer Slide Time: 27:06)



❑ **Fixed effect model: within effect**

- Accomplished by xtreg command with option fe.
- xtreg HEXP HC3 DALE , fe

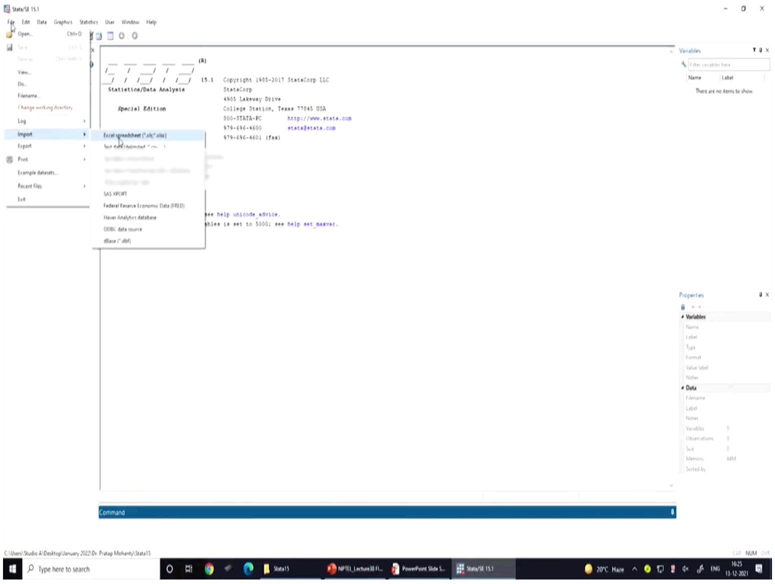
Fixed Effect option

Note!
Add the robust option to control for heteroscedasticity.

- xtreg HEXP HC3 DALE , robust fe

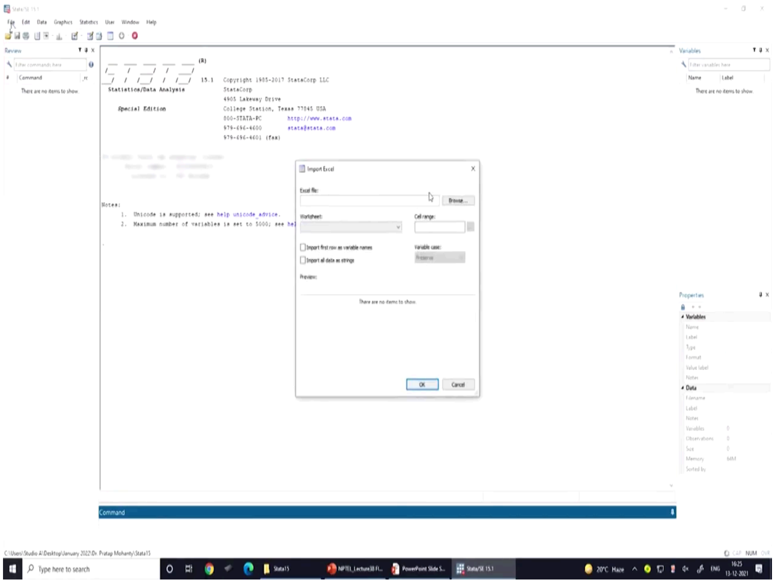
Now, let us even in the next one it will be within effect also would be also captured. So, let us go for it and then clarify.

(Refer Slide Time: 27:14)



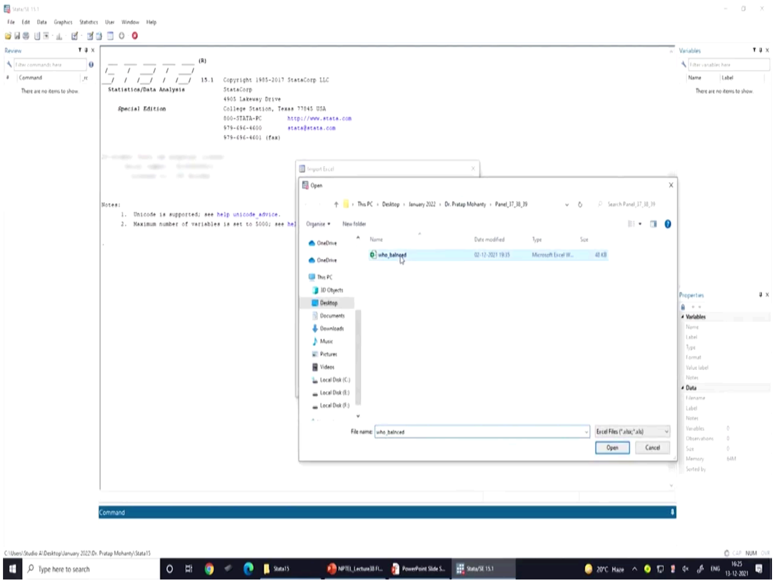
```
15.1   Copyright 1985-2017 StataCorp LLC
Statistics/Data Analysis      StataCorp
                              4905 Lakeway Drive
     Special Edition          College Station, Texas 77845 USA
                              800-STATA-PC        http://www.stata.com
                              979-696-4600        stata@stata.com
                              979-696-4601 (fax)
```

(Refer Slide Time: 27:19)



```
15.1   Copyright 1985-2017 StataCorp LLC
Statistics/Data Analysis      StataCorp
                              4905 Lakeway Drive
     Special Edition          College Station, Texas 77845 USA
                              800-STATA-PC        http://www.stata.com
                              979-696-4600        stata@stata.com
                              979-696-4601 (fax)

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see hel
```

(Refer Slide Time: 27:20)



```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see hel
```

Open

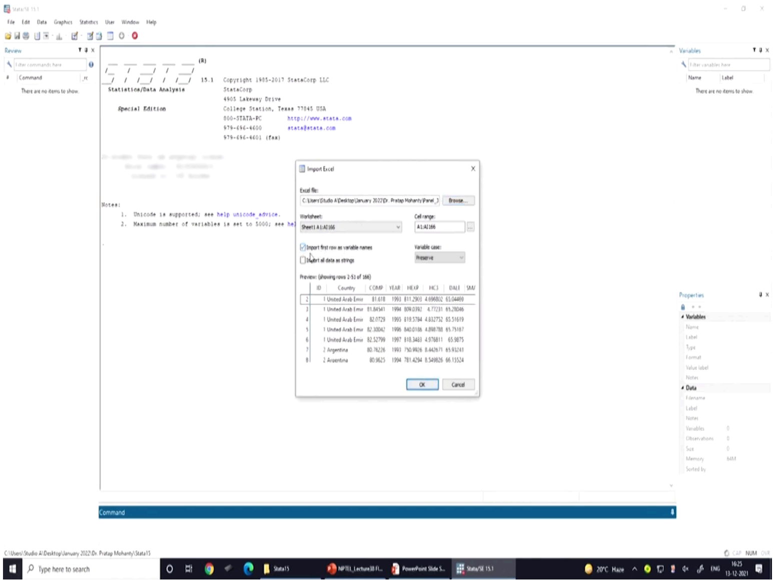
This PC › Desktop › January 2022 › Dr. Pratap Mohanty › Panel_37_38_39

Name	Date modified	Type	Size
who_balnced	02-12-2021 19:35	Microsoft Excel W...	40 KB

File name: who_balnced

Excel Files (*.xlsx;*.xls)

(Refer Slide Time: 27:24)



Import Excel

Excel file: C:\Users\Studio A\Desktop\January 2022\Dr. Pratap Mohanty\Panel_3

Worksheet: Sheet1 A1:AI166

Cell range: A1:AI166

☑ Import first row as variable names

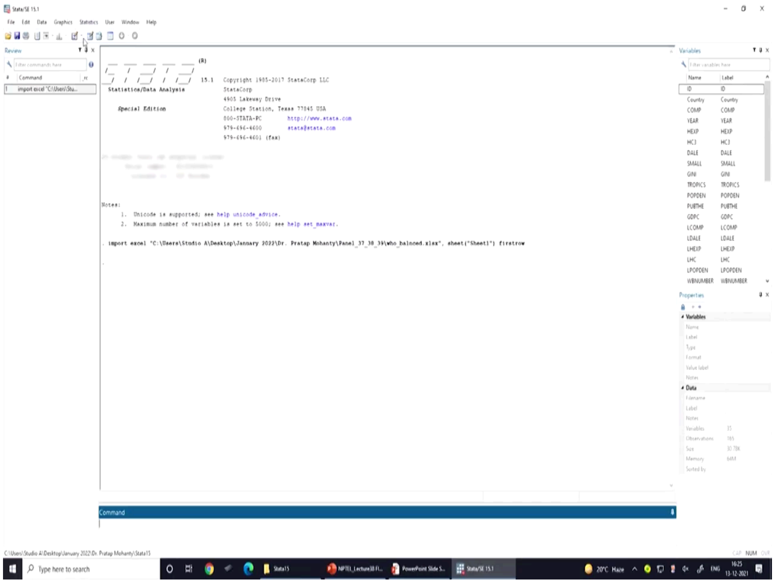
☐ Import all data as strings

Variable case: Preserve

Preview: (showing rows 2-51 of 166)

	ID	Country	COMP	YEAR	HEXP	HC3	DALE	SM/
2	1	United Arab Emir	81.618	1993	811.2903	4.696802	65.04469	
3	1	United Arab Emir	81.84541	1994	809.0392	4.77231	65.28046	
4	1	United Arab Emir	82.0729	1995	819.5784	4.832752	65.51619	
5	1	United Arab Emir	82.30042	1996	840.0186	4.898788	65.75187	
6	1	United Arab Emir	82.52799	1997	818.3483	4.976811	65.9875	
7	2	Argentina	80.78226	1993	750.9926	8.442671	65.93241	
8	2	Argentina	80.9625	1994	781.4294	8.549826	66.15524	

(Refer Slide Time: 27:28)


Stata/SE 15.1
File Edit Data Graphics Statistics User Window Help
Review
Command
import excel "C:\Users\Stu...
Variables
Properties
Command

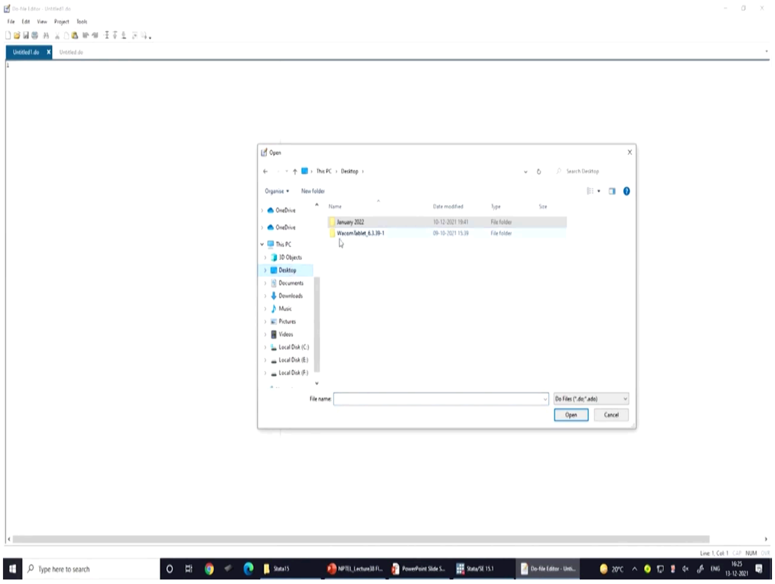
```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

. import excel "C:\Users\Studio A\Desktop\January 2022\Dr. Pratap Mohanty\Panel_37_38_39\who_balnced.xlsx", sheet("Sheet1") firstrow

.
```

Name	Label
ID	ID
Country	Country
COMP	COMP
YEAR	YEAR
HEXP	HEXP
HC3	HC3
DALE	DALE
SMALL	SMALL
GINI	GINI
TROPICS	TROPICS
POPDEN	POPDEN
PUBTHE	PUBTHE
GDPC	GDPC
LCOMP	LCOMP
LDALE	LDALE
LHEXP	LHEXP
LHC	LHC
LPOPDEN	LPOPDEN
WBNUMBER	WBNUMBER

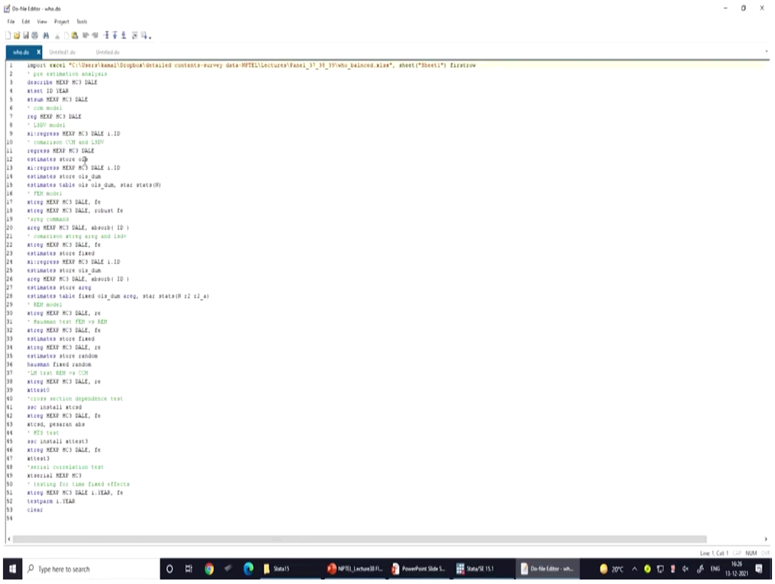
C:\Users\Studio A\Desktop\January 2022\Dr. Pratap Mohanty\Stata15

Here is the data being loaded on your screen and we are taking the Excel data provided by the WHO, example data set. It is balanced data as per the WHO, we will also test that as well. We recognize its variable ID variable as the first row.

(Refer Slide Time: 27:35)



(Refer Slide Time: 27:51)

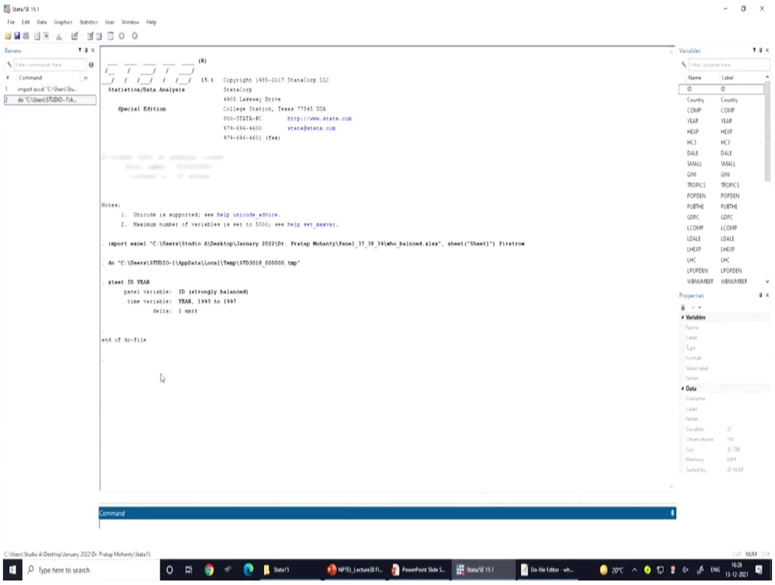


```
import excel "C:\Users\kamal\Dropbox\detailed contents-survey data-NPTEL\Lectures\Panel_37_38_39\who_balnced.xlsx", sheet("Sheet1") firstrow
* pre estimation analysis
describe MEXP MC3 DALE
xtset ID YEAR
xtsum MEXP MC3 DALE
* ccm model
reg MEXP MC3 DALE
* LSDV model
xi:regress MEXP MC3 DALE i.ID
* comarison CCM and LSDV
regress MEXP MC3 DALE
estimates store ols
xi:regress MEXP MC3 DALE i.ID
estimates store ols_dum
estimates table ols ols_dum, star stats(N)
* FEM model
xtreg MEXP MC3 DALE, fe
xtreg MEXP MC3 DALE, robust fe
*areg command
areg MEXP MC3 DALE, absorb( ID )
* comarison xtreg areg and lsdv
xtreg MEXP MC3 DALE, fe
estimates store fixed
xi:regress MEXP MC3 DALE i.ID
estimates store ols_dum
areg MEXP MC3 DALE, absorb( ID )
estimates store areg
estimates table fixed ols_dum areg, star stats(N r2 r2_a)
* REM model
xtreg MEXP MC3 DALE, re
* Hausman test FEM vs REM
xtreg MEXP MC3 DALE, fe
estimates store fixed
xtreg MEXP MC3 DALE, re
estimates store random
hausman fixed random
*LM test REM vs CCM
xtreg MEXP MC3 DALE, re
xttest0
*cross section dependence test
ssc install xtcsd
xtreg MEXP MC3 DALE, fe
xtcsd, pesaran abs
* MT3 test
ssc install xttest3
xtreg MEXP MC3 DALE, fe
xttest3
*serial correlation test
xtserial MEXP MC3
* testing for time fixed effects
xtreg MEXP MC3 DALE i.YEAR, fe
testparm i.YEAR
clear
```

Now, we will also show you the do file that is considered for the analysis, and in the do file, we will be experimenting with the commands and then give you the direction. Now, through the isd through the WHO dataset, we have already explained to you about the CCM model.

CCM model I said that how simple OLS is drawn, but once again we will read this data with stata, xtset should be made first because xtset we will recognize your data as strongly balanced you can see that. And the data is from 93 to 97,

(Refer Slide Time: 28:15)

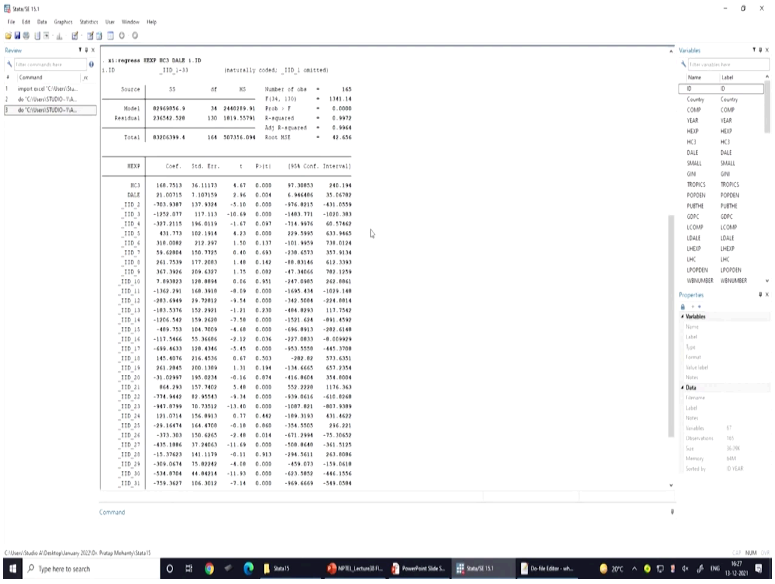


```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

. import excel "C:\Users\Studio A\Desktop\January 2022\Dr. Pratap Mohanty\Panel_37_38_39\who_balnced.xlsx", sheet("Sheet1") firstrow

. do "C:\Users\STUDIO~1\AppData\Local\Temp\STD3018_000000.tmp"

. xtset ID YEAR
       panel variable:  ID (strongly balanced)
        time variable:  YEAR, 1993 to 1997
                delta:  1 unit

.
end of do-file

.
```

Name	Label
ID	ID
Country	Country
COMP	COMP
YEAR	YEAR
HEXP	HEXP
MC3	MC3
DALE	DALE
SMALL	SMALL
GINI	GINI
TROPICS	TROPICS
POPDEN	POPDEN
PUBTHE	PUBTHE
GDPC	GDPC
LCOMP	LCOMP
LDALE	LDALE
LHEXP	LHEXP
LHC	LHC
LPOPDEN	LPOPDEN
WBNUMBER	WBNUMBER

Now, we will be also running the model as per this need regression. We have already made the results earlier for CCM, and now we are clarifying for LSDV model.

(Refer Slide Time: 28:43)



```
. xi:regress HEXP HC3 DALE i.ID
i.ID              _IID_1-33          (naturally coded; _IID_1 omitted)

      Source |       SS           df       MS      Number of obs   =       165
-------------+----------------------------------   F(34, 130)      =   1341.14
       Model |  82969856.9        34  2440289.91   Prob > F        =    0.0000
    Residual |  236542.528       130  1819.55791   R-squared       =    0.9972
-------------+----------------------------------   Adj R-squared   =    0.9964
       Total |  83206399.4       164  507356.094   Root MSE        =    42.656

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   168.7513   36.11173     4.67   0.000     97.30853     240.194
        DALE |   21.00715   7.107159     2.96   0.004     6.946486    35.06782
      _IID_2 |  -703.9387   137.9324    -5.10   0.000    -976.8215   -431.0559
      _IID_3 |  -1252.077    117.113   -10.69   0.000    -1483.771   -1020.383
      _IID_4 |  -327.2115   196.0119    -1.67   0.097    -714.9976    60.57462
      _IID_5 |    431.773   102.1914     4.23   0.000     229.5995    633.9465
      _IID_6 |   318.0082    212.297     1.50   0.137    -101.9959    738.0124
      _IID_7 |   59.62804   150.7725     0.40   0.693    -238.6573    357.9134
      _IID_8 |   261.7539   177.2083     1.48   0.142    -88.83146    612.3393
      _IID_9 |   367.3926   209.6327     1.75   0.082    -47.34066    782.1259
     _IID_10 |   7.893823   128.8894     0.06   0.951    -247.0985    262.8861
     _IID_11 |  -1362.291   168.3918    -8.09   0.000    -1695.434   -1029.148
     _IID_12 |  -283.6949   29.72812    -9.54   0.000    -342.5084   -224.8814
     _IID_13 |  -183.5376   152.2921    -1.21   0.230    -484.8293    117.7542
     _IID_14 |  -1206.542   159.2628    -7.58   0.000    -1521.624   -891.4592
     _IID_15 |   -489.753   104.7009    -4.68   0.000    -696.8913   -282.6148
     _IID_16 |  -117.5466   55.36686    -2.12   0.036    -227.0833   -8.009929
     _IID_17 |  -699.4633   128.4346    -5.45   0.000    -953.5558   -445.3708
     _IID_18 |   145.4076   216.4536     0.67   0.503      -282.82    573.6351
     _IID_19 |   261.2845   200.1309     1.31   0.194    -134.6665    657.2354
     _IID_20 |  -31.02997   195.0234    -0.16   0.874    -416.8684    354.8084
     _IID_21 |    864.293   157.7402     5.48   0.000     552.2228    1176.363
     _IID_22 |  -774.9442   82.95543    -9.34   0.000    -939.0616   -610.8268
     _IID_23 |  -947.8799   70.73512   -13.40   0.000    -1087.821   -807.9389
     _IID_24 |   121.0714   156.8913     0.77   0.442    -189.3193    431.4622
     _IID_25 |  -29.16474   164.4788    -0.18   0.860    -354.5505     296.221
     _IID_26 |   -373.303   150.6265    -2.48   0.014    -671.2994   -75.30652
     _IID_27 |  -435.1886   37.24063   -11.69   0.000    -508.8648   -361.5125
     _IID_28 |  -15.37623   141.1179    -0.11   0.913    -294.5611    263.8086
     _IID_29 |  -309.0674   75.82242    -4.08   0.000     -459.073   -159.0618
     _IID_30 |  -534.8704   44.84214   -11.93   0.000    -623.5852   -446.1556
     _IID_31 |  -759.3627   106.3012    -7.14   0.000    -969.6669   -549.0584
```

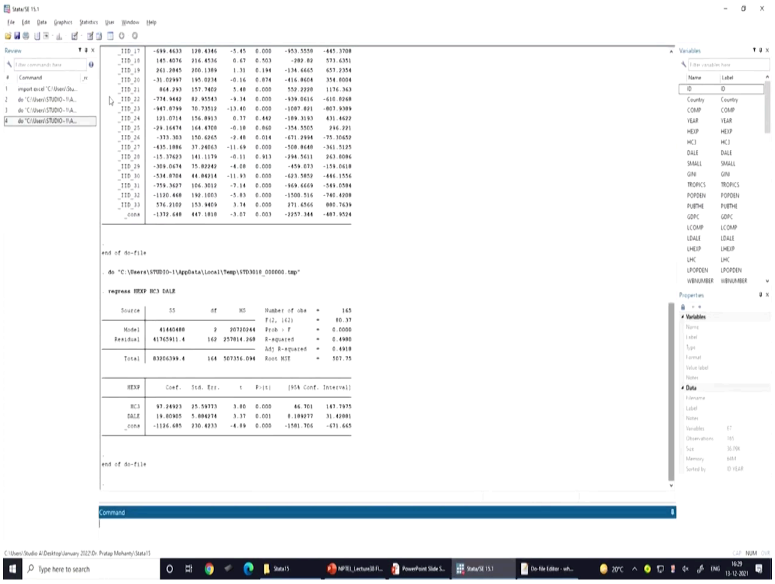
So, LSDV, show the LSDV model is drawn on your screen. You can see how many the dummy variables components are generated, for them to replace the alpha term, so, the number of degrees of freedom is accordingly reduced as I already mentioned, alright.

So, that is one of the wrong disadvantages of this data set. So, dummy variables, the way we explain there are so many dummy variables created, you can easily see this, 33 dummies are there, 32 dummies are therefrom, one is the base category then other dummies are there,

So, the constant term is also there. Show N minus 1 dummy has to be created as I already mentioned since 34 is the N in the model, So, minus 1 is created 30 32 dummies are actually created; the show. And other two so other two are as well there other two variables are also mentioned.

And that is HC3 and DALE. So, all like IID 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 all till 30, yes 33 is mentioned, And in the next one, we will be operating with the next command that is we are we could have drawn the CCM as well. CCM is here,

(Refer Slide Time: 30:45)

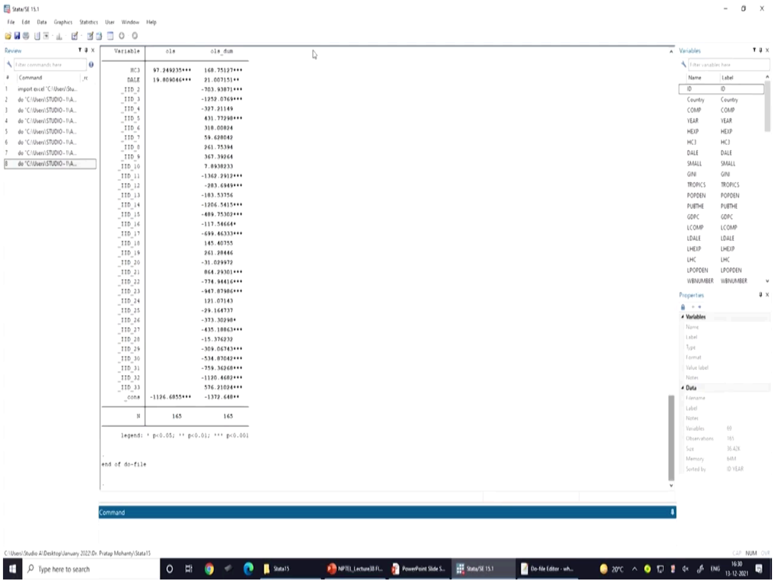


```
       _IID_17 |  -699.4633   128.4346    -5.45   0.000    -953.5558   -445.3708
       _IID_18 |   145.4076   216.4536     0.67   0.503     -282.82    573.6351
       _IID_19 |   261.2045   200.1309     1.31   0.194   -134.6665    657.2354
       _IID_20 |  -31.02997   195.0234    -0.16   0.874   -416.0604    354.0004
       _IID_21 |    864.293   157.7402     5.48   0.000    552.2228    1176.363
       _IID_22 |  -774.9442   82.95543    -9.34   0.000   -939.0616   -610.8268
       _IID_23 |  -947.0799   70.73512   -13.40   0.000   -1087.021   -807.9389
       _IID_24 |   121.0714   156.8913     0.77   0.442   -189.3193    431.4622
       _IID_25 |  -29.16474   164.4708    -0.18   0.860   -354.5505     296.221
       _IID_26 |   -373.303   150.6265    -2.48   0.014   -671.2994   -75.30652
       _IID_27 |  -435.1886   37.24063   -11.69   0.000   -508.8648   -361.5125
       _IID_28 |  -15.37623   141.1179    -0.11   0.913   -294.5611    263.8086
       _IID_29 |  -309.0674   75.82242    -4.08   0.000    -459.073   -159.0618
       _IID_30 |  -534.8704   44.84214   -11.93   0.000   -623.5852   -446.1556
       _IID_31 |  -759.3627   106.3012    -7.14   0.000   -969.6669   -549.0584
       _IID_32 |  -1120.468   192.1003    -5.83   0.000   -1500.516   -740.4208
       _IID_33 |   576.2102   153.9409     3.74   0.000    271.6566    880.7639
         _cons |  -1372.648   447.1818    -3.07   0.003   -2257.344   -487.9524
------------------------------------------------------------------------------

.
end of do-file

. do "C:\Users\STUDIO-1\AppData\Local\Temp\STD3010_000000.tmp"

. regress HEXP HC3 DALE

      Source |       SS           df       MS      Number of obs   =       165
-------------+----------------------------------   F(2, 162)       =     80.37
       Model |    41440488         2    20720244   Prob > F        =    0.0000
    Residual |  41765911.4       162  257814.268   R-squared       =    0.4980
-------------+----------------------------------   Adj R-squared   =    0.4918
       Total |  83206399.4       164  507356.094   Root MSE        =    507.75

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   97.24923   25.59773     3.80   0.000       46.701    147.7975
        DALE |   19.80905   5.884274     3.37   0.001     8.189277    31.42881
       _cons |  -1126.685   230.4233    -4.89   0.000    -1581.706    -671.665
------------------------------------------------------------------------------

.
end of do-file

.
```

So, just to regress the model we are trying to compare the equation and we will store this so that that is going to be helpful for us to compare, Now, this is being stored in our memory. Now, we will give the conditional command with i dot command the way we do since we have considered the dummy variables model.

Now, the result is displayed. So, now, we are also storing the estimates of this dummy with the i command, i dot command that we have given i dot id is given. So, it is basically comparing with the first categories of the dummy. Now, this is going to give us what we have already stored we can easily compare. Now, estimates will compare each of the tables that is OLS, OLS dummy, and with it a star and at what level this varies we can easily do it and the result is on your screen,

(Refer Slide Time: 31:55)



```
    Variable        ols        ols_dum
                                      
         HC3   97.249235***  160.75127***
        DALE   19.809046***  21.007151**
      _IID_2                -703.93071***
      _IID_3               -1252.0769***
      _IID_4                -327.21149
      _IID_5                 431.77298***
      _IID_6                 318.00824
      _IID_7                 59.620842
      _IID_8                 261.75394
      _IID_9                 367.39264
     _IID_10                 7.8938233
     _IID_11               -1362.2912***
     _IID_12                -283.6949***
     _IID_13                -103.53756
     _IID_14               -1206.5415***
     _IID_15                -489.75302***
     _IID_16                -117.54664*
     _IID_17                -699.46333***
     _IID_18                 145.40755
     _IID_19                 261.20446
     _IID_20                -31.029972
     _IID_21                 864.29301***
     _IID_22                -774.94416***
     _IID_23                -947.07906***
     _IID_24                 121.07143
     _IID_25                -29.164737
     _IID_26                -373.30298*
     _IID_27                -435.18863***
     _IID_28                -15.376232
     _IID_29                -309.06743***
     _IID_30                -534.87042***
     _IID_31                -759.36268***
     _IID_32               -1120.4682***
     _IID_33                 576.21024***
       _cons  -1126.6855***  -1372.648**
                                      
           N        165          165
                                      
      legend: * p<0.05; ** p<0.01; *** p<0.001

.
end of do-file

.
```

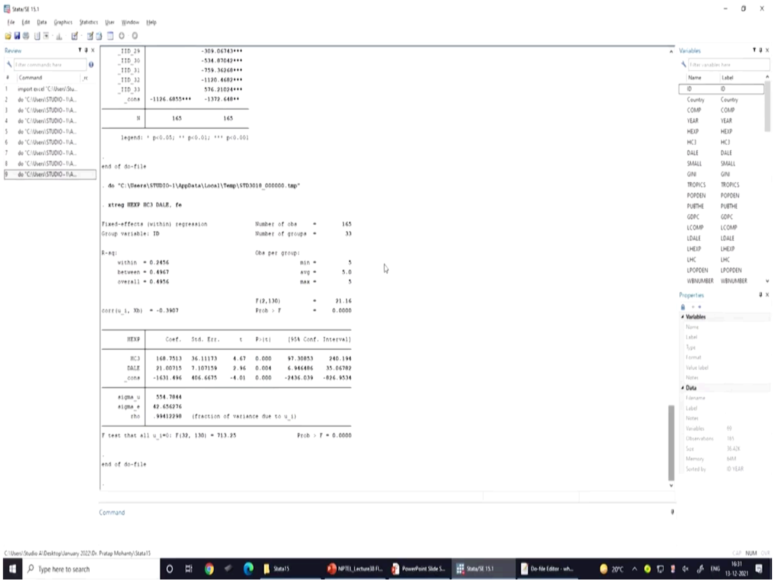
Now, OLS and the dummy variables model, so how this is important and how each of these is different, you can see that in the normal OLS or with the CCM model. It has only two coefficients and estimating, it is similar to that of the OLS model. But in the case of dummy one, it estimates so many coefficients.

Though we have run the OLS model, we can easily identify each of the effects differently as compared to the base category. And disadvantages of this we have already discussed therefore, we are going to explain the fixed effect model. So, the fixed-effect model we are we are now explaining.

In that case, we need to identify the stata per the correct specification with fe. We need to understand whether fe is specified or not. But there are some tests to understand whether it is a fixed effect or random effect that we will do in our next class,

So, the exact Horseman Test how that is significantly identifying or not that aspect we will discuss. Let us assume that it has certain fixed effects or within effects and time-variant effects. So, that is what we have assumed, and based on that we have run the regression.

(Refer Slide Time: 33:25)



```
     _IID_29                   -309.06743***
     _IID_30                   -534.07042***
     _IID_31                   -759.36260***
     _IID_32                  -1120.4602***
     _IID_33                    576.21024***
       _cons  -1126.6055***   -1372.640**
--------------------------------------------
           N        165            165
--------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001

.
end of do-file

. do "C:\Users\STUDIO~1\AppData\Local\Temp\STD3010_000000.tmp"

. xtreg HEXP HC3 DALE, fe

Fixed-effects (within) regression               Number of obs     =        165
Group variable: ID                              Number of groups  =         33

R-sq:                                           Obs per group:
     within  = 0.2456                                         min =          5
     between = 0.4967                                         avg =        5.0
     overall = 0.4956                                         max =          5

                                                F(2,130)          =      21.16
corr(u_i, Xb)  = -0.3907                        Prob > F          =     0.0000

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   160.7513   36.11173     4.67   0.000     97.30853     240.194
        DALE |   21.00715   7.107159     2.96   0.004     6.946406    35.06702
       _cons |  -1631.496   406.6675    -4.01   0.000    -2436.039   -826.9534
-------------+----------------------------------------------------------------
     sigma_u |  554.7044
     sigma_e |  42.656276
         rho |  .99412298   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(32, 130) = 713.25                   Prob > F = 0.0000

.
end of do-file

.
```

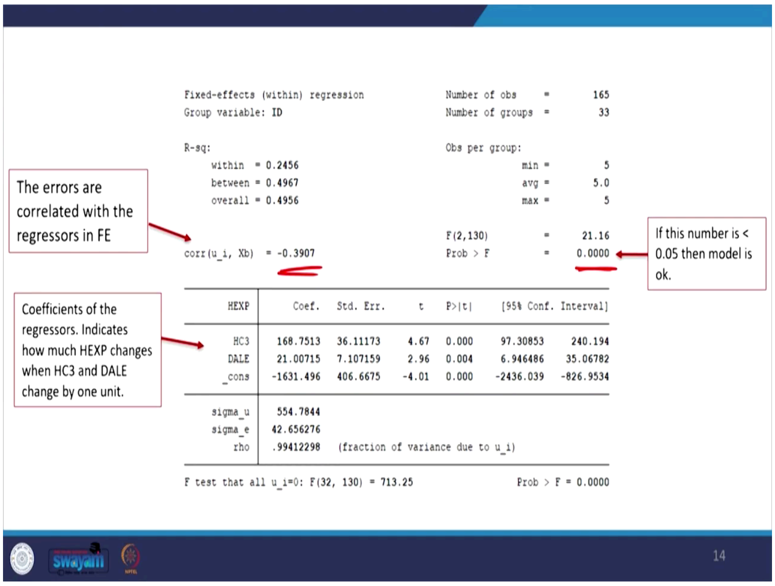
So, in the fixed-effect model, the correlation is more important. The correlation between the error term and the explanatory variable is there, it is not 0, So, the correlation is identifying the fact that it is, in fact, a model where the alpha i content, alpha actually content is correlated with the explanatory variables,

So, let me just explain to you through the model ones again, through the PPT ones again, then I will come back to the discussion. We have covered all those aspects. These are the command you have to follow the way we did it, we have kept capturing everything in the slide and that will help you to read between the results.

Now, in the within the model, within effect model we have already clarified what is called the LSDV model and what is called within effect model. And xtreg, r e g, xtreg command with finally, at ending with command fe will capture the fixed effect. And adding the robust option is going to control the heteroscedasticity.

So, usually, there are heteroscedasticity in the model, as I already told you. there are it is not avoiding the problem of heteroscedasticity. In order to control the heteroscedasticity, we usually take the command called robust fe, So, xtreg in at the end robust fe is going to control certain heteroscedasticity to run the fixed effect model. This is what it will lo like. And I will come back to it. We have already done this.

(Refer Slide Time: 35:31)



```
Fixed-effects (within) regression               Number of obs      =       165
Group variable: ID                              Number of groups   =        33

R-sq:                                           Obs per group:
     within  = 0.2456                                         min =          5
     between = 0.4967                                         avg =        5.0
     overall = 0.4956                                         max =          5

                                                F(2,130)           =     21.16
corr(u_i, Xb)  = -0.3907                        Prob > F           =    0.0000

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   168.7513   36.11173     4.67   0.000     97.30853     240.194
        DALE |   21.00715   7.107159     2.96   0.004     6.946486    35.06782
       _cons |  -1631.496   406.6675    -4.01   0.000    -2436.039   -826.9534
-------------+----------------------------------------------------------------
     sigma_u |   554.7844
     sigma_e |  42.656276
         rho |  .99412298   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(32, 130) = 713.25                   Prob > F = 0.0000
```

The errors are correlated with the regressors in FE

If this number is < 0.05 then model is ok.

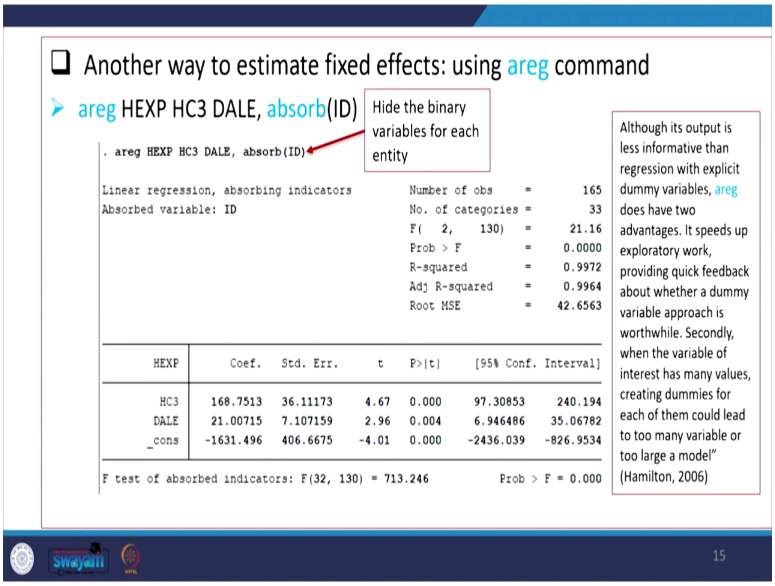
Coefficients of the regressors. Indicates how much HEXP changes when HC3 and DALE change by one unit.

This is the one we derived and I am interpreting it through your PPT once again. And in the PPT, the first aspect is that the errors are correlated with the regressions in the fixed-effect model, And that is due to the fixed effect component, captured in what is called the individual effects or the constant term.

And that is why it has certain correlation with the error term as well. So, it is nonzero, therefore, there are some problems with the model estimation. We cannot just run the run a normal ordinarily square regression. Now, the second one is if this number is the probability the significance value overall significance of the model, of this, is less than 0.05 then the model is fine,

Then, the coefficients have been interpreted the way we usually do it. This coefficient indicates like the household the health expenditure per capita changes when the disability-adjusted life expectations and the education etcetera is changed by one unit, So, whether they are significant or not etcetera can also be interpreted accordingly.

(Refer Slide Time: 36:50)



❑ Another way to estimate fixed effects: using areg command

➢ areg HEXP HC3 DALE, absorb(ID)

Hide the binary variables for each entity

```
. areg HEXP HC3 DALE, absorb(ID)

Linear regression, absorbing indicators         Number of obs      =       165
Absorbed variable: ID                           No. of categories =         33
                                                F(   2,     130)  =      21.16
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9972
                                                Adj R-squared     =     0.9964
                                                Root MSE          =    42.6563

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   168.7513   36.11173     4.67   0.000     97.30853     240.194
        DALE |   21.00715   7.107159     2.96   0.004     6.946486    35.06782
       _cons |  -1631.496   406.6675    -4.01   0.000    -2436.039   -826.9534
------------------------------------------------------------------------------
F test of absorbed indicators: F(32, 130) = 713.246           Prob > F = 0.000
```

Although its output is less informative than regression with explicit dummy variables, areg does have two advantages. It speeds up exploratory work, providing quick feedback about whether a dummy variable approach is worthwhile. Secondly, when the variable of interest has many values, creating dummies for each of them could lead to too many variable or too large a model" (Hamilton, 2006)

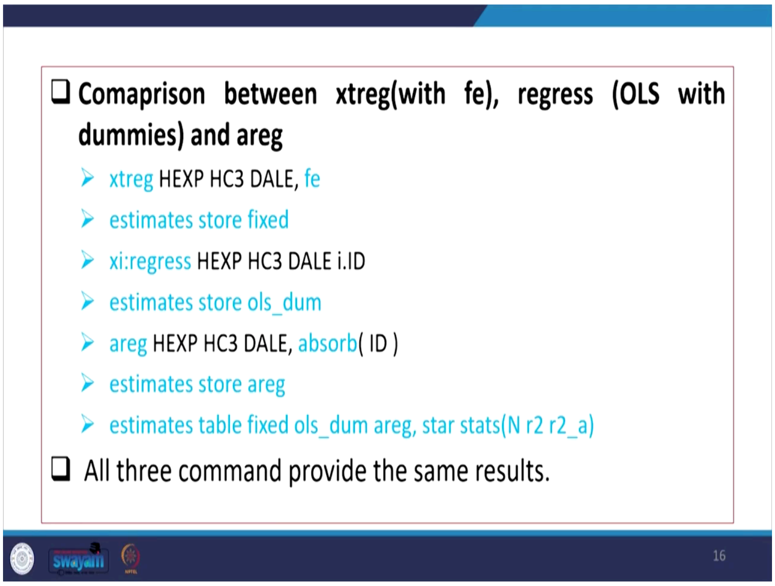
15

Another way to estimate the fixed effect using areg command. areg command actually this is going to observe the problems and it is going to hide the binary variables for each entity, So, the since binary variables are there that is going to observe it correctly. Although its output is less informative than regression with explicit dummy variables since explicit dummy variables are there areg command is the base fit and this does have two advantages.

It speeds the speeds of explanatory work providing quick feedback about whether a dummy variable approach is worthwhile. The second one is called when the second one is on when the variable of interests has many values creating dummies for each of them could lead to too many variables or too large a model in that case as per the suggestions Hamilton that you please go through the areg command, that observes the problems with the too many indicator variables,

So, that is also going to give you the result. The commands are already on your screen. You can apply on the same data, I am sure you will get it,

(Refer Slide Time: 38:15)



❑ Comaprison between xtreg(with fe), regress (OLS with dummies) and areg

- xtreg HEXP HC3 DALE, fe
- estimates store fixed
- xi:regress HEXP HC3 DALE i.ID
- estimates store ols_dum
- areg HEXP HC3 DALE, absorb(ID)
- estimates store areg
- estimates table fixed ols_dum areg, star stats(N r2 r2_a)

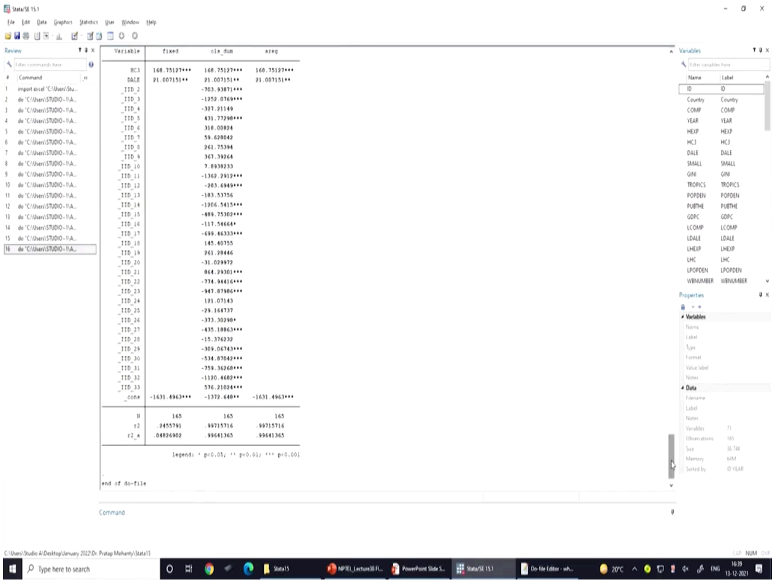
❑ All three command provide the same results.

So, that we can have a comparison. Let me also experiment with it and going to interpret it accordingly, . Now, we are giving you the comparison, we will draw on each of them. Then, so the fixed effect result is being derived at this moment and we will also store this, . We will also store this. So, with the name fixed, and now once again we will draw the LSDV approach with this conditional command with i dot id. And again, we will store with an OLS dummy that is LSDV regression.

Then, we will also compare with the try to observe the identification variable or the dummy variables. Then a areg is the command and we will also store with areg. And now we can after doing so, we can estimate a table with the comparison of each of them. So, each of them now is being stored and now you can each of the they is estimated. Now, you can just see the differences.

So, we have all derived. The results are derived on your screen. So, let us see once,

(Refer Slide Time: 39:49)

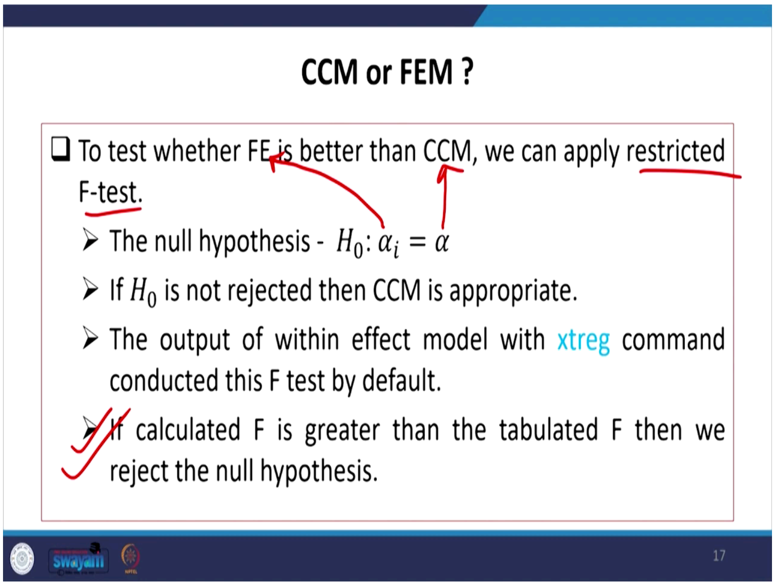


```
    Variable        fixed         ols_dum          areg
-------------------------------------------------------
         HC3   160.75127***   160.75127***   160.75127***
        DALE   21.007151**    21.007151**    21.007151**
      _IID_2                 -703.93071***
      _IID_3                 -1252.0769***
      _IID_4                 -327.21149
      _IID_5                  431.77298***
      _IID_6                  310.00024
      _IID_7                  59.620042
      _IID_8                  261.75394
      _IID_9                  367.39264
     _IID_10                  7.0930233
     _IID_11                 -1362.2912***
     _IID_12                  -203.6949***
     _IID_13                 -103.53756
     _IID_14                 -1206.5415***
     _IID_15                 -409.75302***
     _IID_16                 -117.54664*
     _IID_17                 -699.46333***
     _IID_18                  145.40755
     _IID_19                  261.20446
     _IID_20                 -31.029972
     _IID_21                  864.29301***
     _IID_22                 -774.94416***
     _IID_23                 -947.07906***
     _IID_24                  121.07143
     _IID_25                 -29.164737
     _IID_26                 -373.30298*
     _IID_27                 -435.10063***
     _IID_28                 -15.376232
     _IID_29                 -309.06743***
     _IID_30                 -534.07042***
     _IID_31                 -759.36260***
     _IID_32                 -1120.4602***
     _IID_33                  576.21024***
       _cons  -1631.4963***   -1372.640**   -1631.4963***
-------------------------------------------------------
           N         165           165           165
          r2     .2455791      .99715716     .99715716
        r2_a     .04026902     .99641365     .99641365
-------------------------------------------------------
              legend: * p<0.05; ** p<0.01; *** p<0.001

.
end of do-file
```

So, all 3 are in fact, presented on your screen. This is I think you might have understood the way we proceeded. All 3 are stored. The first one is fixed-effect model the fe g, fe command is given, then the LSDV model is here. Now, LSDV is also reduced or is interpreted in a better way through areg, that has been made.

And now you can see you will get the same result, through the areg. areg and fixed effect model are going to give you when you have already captured the dummy variables problems, So, any of the models you can do it, run it and the any of these two models. You can do it and estimate it, fine. These are all for this lecture. We probably, we have if any things are there we can also explain you otherwise we will; yes.

(Refer Slide Time: 40:57)



CCM or FEM ?

- ❑ To test whether FE is better than CCM, we can apply restricted F-test.
  - ➢ The null hypothesis - $H_0: \alpha_i = \alpha$
  - ➢ If H_0 is not rejected then CCM is appropriate.
  - ➢ The output of within effect model with xtreg command conducted this F test by default.
  - ➢ If calculated F is greater than the tabulated F then we reject the null hypothesis.

We can have some comparison as well with the constant common model results, it is not just the 3 within effects of LSDV. In fact, we can also compare the simple regression estimation as well as the fixed effect coefficient as differences. To test whether fe is better than CCM or not we can apply restra icted F test, .

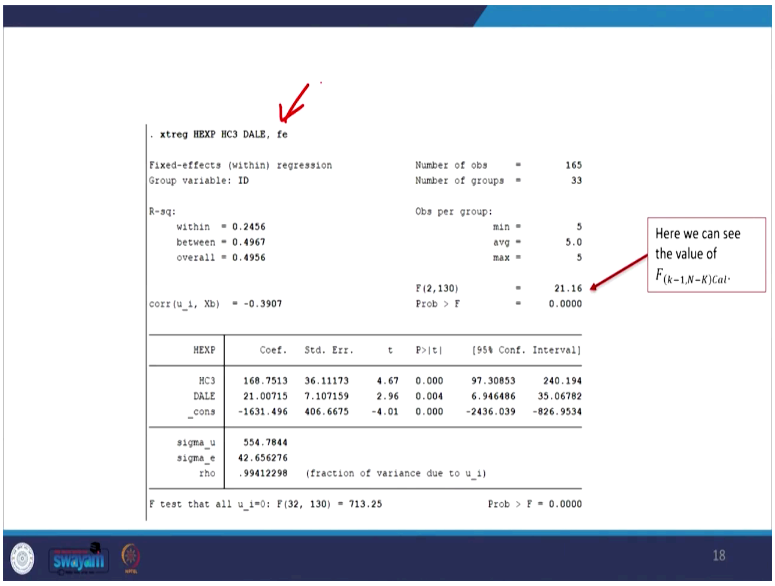
So, the restricted F test is going to give you the direction. The null hypothesis, in this case, is whether alpha i is equal to alpha because alpha is the one in case of CCM, and alpha is the one in the case of FE. So, whether these two are actually equal or not. S, so the null hypothesis is going to be considered as equal. So, our hypothesis is in fact, alpha is equal to alpha.

If the null hypothesis is not rejected, then CCM is appropriate, If both are the same then what is the point of going for a panel regression; if they are different then, of course, you should go for fe test or fixed-effect model. The output of the within effect model with xtreg command a conducted this F test by default,

If the calculated F is greater than the tabulated F, then we will reject the null hypothesis. So, this is the one that is very important; when your calculated F is is different, or greater than that of your tabulated one; the tabulated F value then, of course, that is going to reject the null hypothesis. And fe is going to be considered to be better.

In this case, you can see that here is the calculated value which is 21.16 that is the F value.

(Refer Slide Time: 42:54)

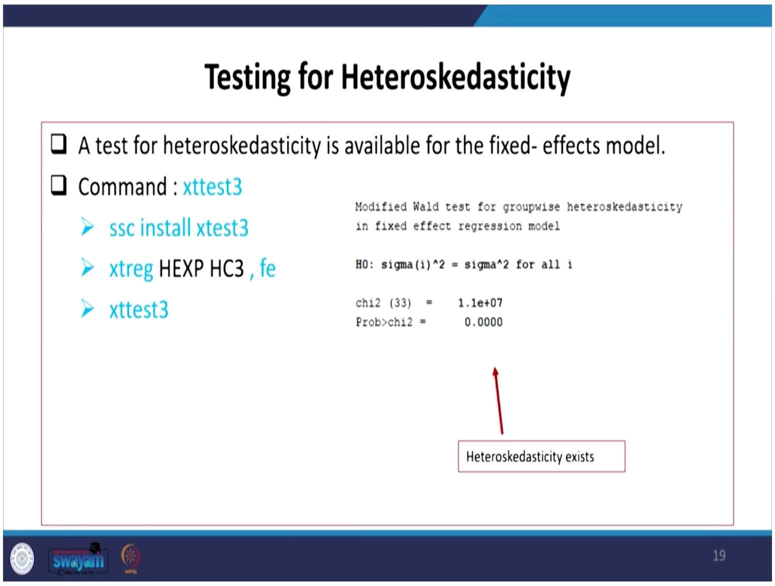


```
. xtreg HEXP HC3 DALE, fe

Fixed-effects (within) regression               Number of obs     =        165
Group variable: ID                              Number of groups  =         33

R-sq:                                           Obs per group:
     within  = 0.2456                                         min =          5
     between = 0.4967                                         avg =        5.0
     overall = 0.4956                                         max =          5

                                                F(2,130)          =      21.16
corr(u_i, Xb)  = -0.3907                        Prob > F          =     0.0000

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   168.7513   36.11173     4.67   0.000     97.30853     240.194
        DALE |   21.00715   7.107159     2.96   0.004     6.946486    35.06782
       _cons |  -1631.496   406.6675    -4.01   0.000    -2436.039   -826.9534
-------------+----------------------------------------------------------------
     sigma_u |  554.7844
     sigma_e |  42.656276
         rho |  .99412298   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(32, 130) = 713.25                   Prob > F = 0.0000
```

Here we can see the value of $F_{(k-1,N-K)Cal}$.

With the xtreg command with fe command, we have given, . If this is this is the one then we will compare with the F value with its degrees of freedom, degrees of freedom with the numerator and denominator values, k minus 1 in the numerator and N minus K in the denominator the calculated value we can compare. If the tabulated value calculated value is different than that of the tabulated value, then we will go for rejecting the null hypothesis,

(Refer Slide Time: 43:25)



Testing for Heteroskedasticity

- ❑ A test for heteroskedasticity is available for the fixed- effects model.
- ❑ Command : xttest3
  - ➢ ssc install xttest3
  - ➢ xtreg HEXP HC3 , fe
  - ➢ xttest3

```
Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (33)  =    1.1e+07
Prob>chi2 =      0.0000
```

Heteroskedasticity exists

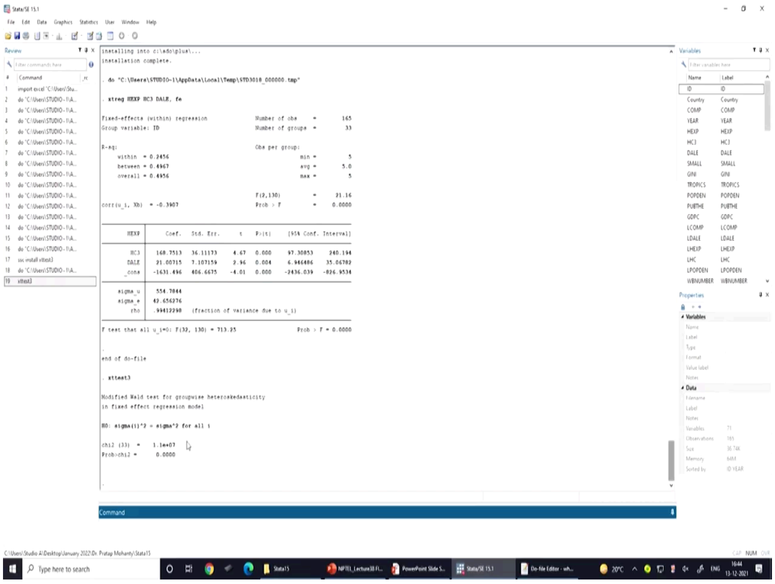
So, that next one is called testing of heteroscedasticity which is also equally important in case of the fixed-effect model. We have already said that it does not avoid the problems of heteroscedasticity. So, a test is quite essential. A test of heteroscedasticity is available for the fixed effect model because of its assumptions, So, the commands we usually give it is xttest3, then we need to actually install it and then run the xttest3 because it is not come up with the version by default.

Then we can run with the xtreg and its fe and that and at the end after the regression, we need to just test xttest3, I will do that here, but I am just interpreting it if that is the case that the assumption is that should be homogeneous, that should not be any heterogeneous.

But if it is significant if the p values are significant, the null hypothesis is that the standard deviation the sigma is actually equal for all is, but if that is significant; that means, they are not equal, so heteroscedasticity exists. So, this is what I am going to show on the screen.

Once, then we can go for it and then explain it, So, we will install that first SSE install xttest3. So, this is being installed, Now, this is and verifying not installed. We can do it then.

(Refer Slide Time: 45:18)



```
installing into c:\ado\plus\...
installation complete.

. do "C:\Users\STUDIO~1\AppData\Local\Temp\STD3018_000000.tmp"

. xtreg HEXP HC3 DALE, fe

Fixed-effects (within) regression               Number of obs     =        165
Group variable: ID                              Number of groups  =         33

R-sq:                                           Obs per group:
     within  = 0.2456                                         min =          5
     between = 0.4967                                         avg =        5.0
     overall = 0.4956                                         max =          5

                                                F(2,130)          =      21.16
corr(u_i, Xb)  = -0.3907                        Prob > F          =     0.0000

------------------------------------------------------------------------------
        HEXP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         HC3 |   160.7513   36.11173     4.67   0.000     97.30853     240.194
        DALE |   21.00715   7.107159     2.96   0.004     6.946406    35.06782
       _cons |  -1631.496   406.6675    -4.01   0.000    -2436.039   -826.9534
-------------+----------------------------------------------------------------
     sigma_u |  554.7044
     sigma_e |  42.656276
         rho |  .99412298   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(32, 130) = 713.25                   Prob > F = 0.0000

.
end of do-file

. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (33)  =    1.1e+07
Prob>chi2 =      0.0000

.
```

Then, we will now run it and now the installation is completed. Now, we are running the fixed effect model in our case. So, we have already run the fixed-effect model and now we are going to test the xttest3 that is for the heteroscedasticity. Now, it is confirmed in the case that our significance level, the sigma values are actually differing that is differing significantly therefore, there exists heteroscedasticity in the model.

So, that is another way of proving that yes we are running the fixed-effect model and it has heteroscedasticity in the model and it is not going by the common constant model, The common constant model is not the right fit when this kind of test are identifying the heteroscedasticity, alright. So, the fixed-effect model is best chosen.

And in the next class, we will discuss choosing between choice between fixed-effect model and the random effect model, and we will also give you various other dimensions of clarifying random effect in the healthcare database. We will be also using IHDS data. So, we have also kept the IHDS data. Your own reference you can also test on your own.

We have only operated through the WHO data set. If you still have some confusion you can also refer to my earlier model, we have used IHDS data. This time we have we are also keeping the same IHDS data for your reference. We have operated with one database, you can operate on your own and find out the difference. The do-file of both the cases has been uploaded and that will be going to be useful for you to run it.

I hope this is how you can get the best understanding out of it. And if still you have difficulties do not hesitate and come back to us in our live session or in our chat session. Our team is quite charged to deal with all your queries.

Thank you.