Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee

Lecture - 32
Common Constant Model in Healthcare

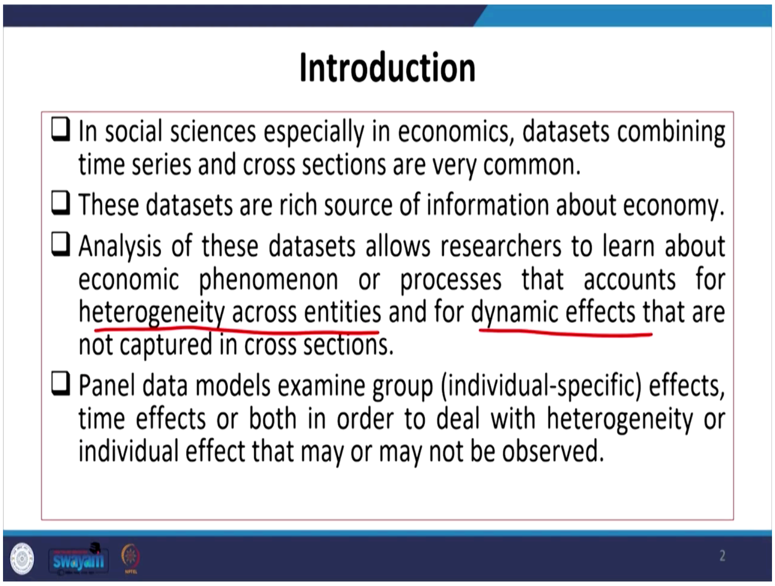
Welcome friends once again to my NPTEL MOOC module on Exploring Healthcare Survey Data. We are on the verge of 7th week of the module; we have already started explaining panel data in healthcare. On this particular lecture we are directed to deal with the different models of panel data and healthcare.

Here we have thought of dealing with one of the very starting points of panel model usually discussed as or called as common constant model. Otherwise called pooled model in panel or in healthcare. So, let us understand to what extent pooled or common constant in a model are different than that of the proper panel model.

This lecture week we are going to give largely on clarifying the concept and with some practical handouts, practical experiments. We will also nurture you with the exact interpretations of these all models and these are highly useful in research specially those are advanced users.

Advanced users in the sense those who are writing their papers and submitting to a good journal usually A category journals. For them I will always suggest to follow this very lecture carefully and then accomplish your paper accordingly. So, here are the background information about this model.

(Refer Slide Time: 02:02)



Introduction

- ❑ In social sciences especially in economics, datasets combining time series and cross sections are very common.
- ❑ These datasets are rich source of information about economy.
- ❑ Analysis of these datasets allows researchers to learn about economic phenomenon or processes that accounts for heterogeneity across entities and for dynamic effects that are not captured in cross sections.
- ❑ Panel data models examine group (individual-specific) effects, time effects or both in order to deal with heterogeneity or individual effect that may or may not be observed.

In social sciences, especially in economics data sets, combining time series and cross sections are usually very prevalent. These data sets are rich source of information regarding the activities of an economy. To analyze these databases, we allow researchers to learn about how economic dimensions or structure accounts for heterogeneity across entities or dynamic effects that are not captured in cross sections.

So, two important points are to be highlight here, one is called heterogeneity across entities. So, it’s not just the cross-sectional heterogeneity, there are heterogeneity due to other dimensions as well. In addition to that we are also adding information about dynamic effects which are not captured in cross sections data.

The panel data models examine group effects, time effects or both in order to deal with heterogeneity or individual effect that may or may not be observed. So, like it captures individual specific effects as well as group effects, both the cases are going to be discussed through this lecture.

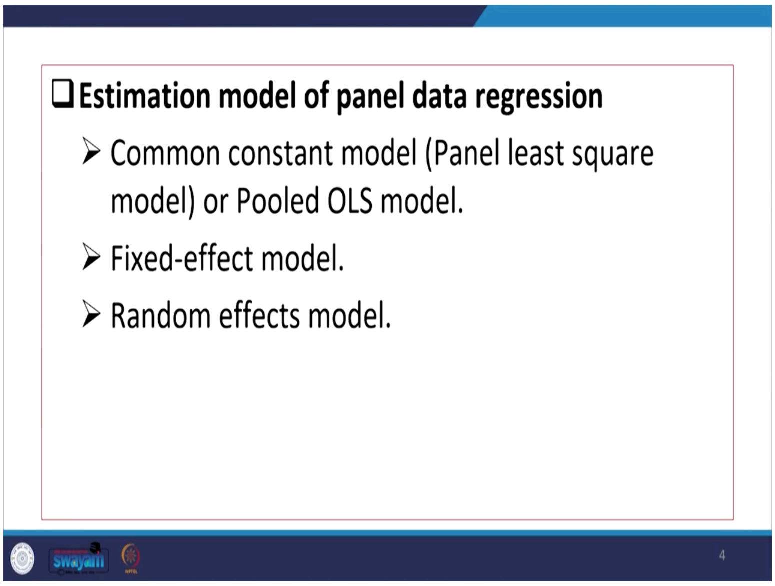
The panel data methods can be divided into two broad categories one is called homogeneous panel data; second one is called heterogeneous panel data. Homogeneous, as the word clarifies this assumes that the model parameters are actually homogeneous. In the sense they are common across individuals like we have seen in ordinarily square method or ordinary regression methods or in a general cross-sectional data.

These are also called pooled models. So, homogeneous panel data are discussed in the context of pooled models as well. Whereas, in the case of heterogeneous panel data models, this allows the model to have different parameters across individuals.

The individuals are actually varying with their predicted values when there are some forms of dynamism, some forms of differences across individuals may be due to the differences across time or across the cross sections.

So, when both are understood they are actually captured in the heterogeneous panel model. Another aspect is called fixed effect and random effect models within the heterogeneous panel model data. These both models actually are example of panel data models, heterogeneous panel data models.

(Refer Slide Time: 05:11)

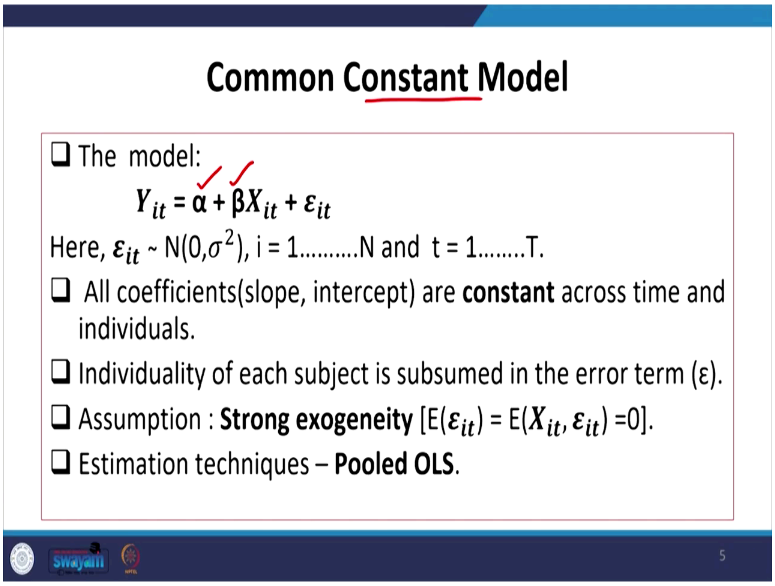


❑ **Estimation model of panel data regression**

- ➢ Common constant model (Panel least square model) or Pooled OLS model.
- ➢ Fixed-effect model.
- ➢ Random effects model.

To estimate the panel data regression, we consider three important models. One is called Common constant model also called Panel least square model. These are also called Pooled OLS regression model. Then another one is called Fixed effect model and the third one is called Random effect model.

(Refer Slide Time: 05:24)



Common Constant Model

- ❑ The model:

$$Y_{it} = \overset{\checkmark}{\alpha} + \overset{\checkmark}{\beta} X_{it} + \varepsilon_{it}$$

Here, $\varepsilon_{it} \sim N(0,\sigma^2)$, i = 1..........N and t = 1........T.

- ❑ All coefficients(slope, intercept) are **constant** across time and individuals.
- ❑ Individuality of each subject is subsumed in the error term (ε).
- ❑ Assumption : **Strong exogeneity** [$E(\varepsilon_{it}) = E(X_{it}, \varepsilon_{it}) = 0$].
- ❑ Estimation techniques – **Pooled OLS**.

Let us start with the common constant model which we could have written as pooled. You might have further raised certain confusion to do that we use the term common constant model as per the book and as per the standard definition or explanation.

So, the word common constant means your coefficient of estimation is going to be constant throughout the model for the particular estimation. Now, you can see which are the common estimate that is the constant estimator and this is the coefficient or slope estimator. Both are actually going to be constant slope as well as intercept are constant across time as well as individuals in such cases it is called common constant model.

Now assumption similar to that of the OLS model is that the error term though that varies across cross section and time. But in such type of data, we simply assume that it is normally distributed and then with standard normal distribution with it is mean as 0 and standard variance is sigma square.

So, here i varies from 1 to N across cross-sections and t varies from 1 to capital T time points. So, the individuality of each subject is; of each subject is subsumed in the error term. So, any sort of individual differences those are actually captured in error term. Therefore, the assumption in this case is that there must be strong heterogeneity. Heterogeneity between the explanatory variable and the error term. So, expected value of the explanatory variable and the error term is equal to 0.

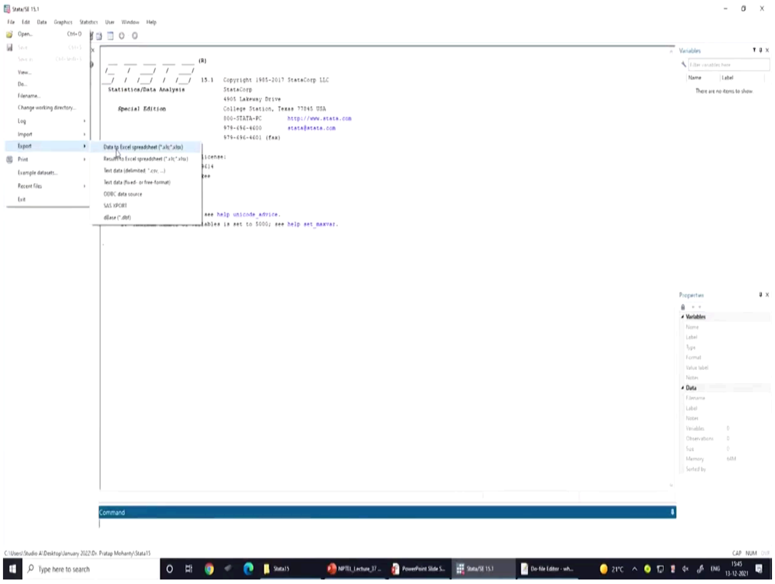
So, the estimation technique we consider is called pooled OLS model because simply each item by varying by time or by varying by cross section. These are simply pooled together and defined with a common model and that is why they are called pooled.

Now, you might have questions on this that how come they are even if pooled there are differences. Our assumption is forcefully making it with no endogeneity, it has strong exogeneity. Therefore, we go for OLS. Even if we do it, we cannot just convince ourselves that it is going to have 0 exogeneity or 0 endogeneity. It has strong exogeneity there are some practical examples we wanted to show it to you and we are going to experiment with it.

Like here we are giving a pool kind of data. We have also cited the data from WHO panel data sets those are available. And we will also make you available with this practical dataset in your portal and all of you could able to find the data for your own experimentation. Along with that we are going to also show UN and run regression results with IHDS panel data. I just want to mention you just give you a note again that if you have still some confusions about forming a panel data through IHDS you need to refer to my earlier module i.e., handling large scale dataset using STATA that is still running at this moment. And you may also follow from YouTube that how panel data could be generated. IHDS was initially not providing the panel data to the public. But nowadays the structured panel data is available for us to work with.

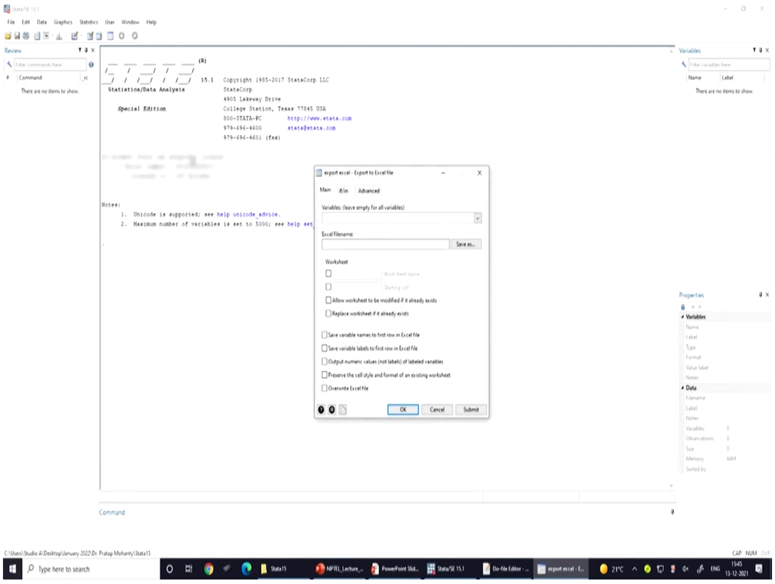
So, both the way we will be emphasising and clarifying your doubt. So, IHDS is the only longitudinal survey large scale survey available in India. There are in fact, two waves of this survey IHDS 1 and IHDS 2. So, let us go for it and explain a bit about it.

(Refer Slide Time: 09:56)


Stata/SE 15.1
File Edit Data Graphics Statistics User Window Help
Open... Ctrl+O
Save Ctrl+S
Save as... Ctrl+Shift+S
View...
Do...
Filename...
Change working directory...
Log
Import
Export
Print
Example datasets...
Recent files
Exit
Data to Excel spreadsheet (*.xls;*.xlsx)
Results to Excel spreadsheet (*.xls;*.xlsx)
Text data (delimited, *.csv, ...)
Text data (fixed- or free-format)
ODBC data source
SAS XPORT
dBase (*.dbf)
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)
```

(Refer Slide Time: 09:58)


Stata/SE 15.1
File Edit Data Graphics Statistics User Window Help
Review
Command
There are no items to show.
Variables
Name Label
There are no items to show.
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.
```

export excel - Export to Excel file

Main | if/in | Advanced

Variables: (leave empty for all variables)

Excel filename: [Save as...]

Worksheet

□ Worksheet name

□ Starting cell

□ Allow worksheet to be modified if it already exists

□ Replace worksheet if it already exists

□ Save variable names to first row in Excel file

□ Save variable labels to first row in Excel file

□ Output numeric values (not labels) of labeled variables

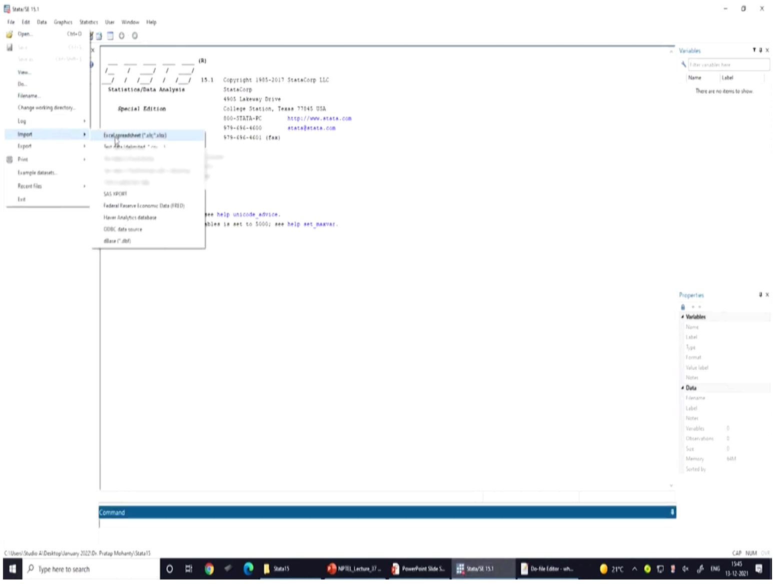
□ Preserve the cell style and format of an existing worksheet

□ Overwrite Excel file

OK | Cancel | Submit

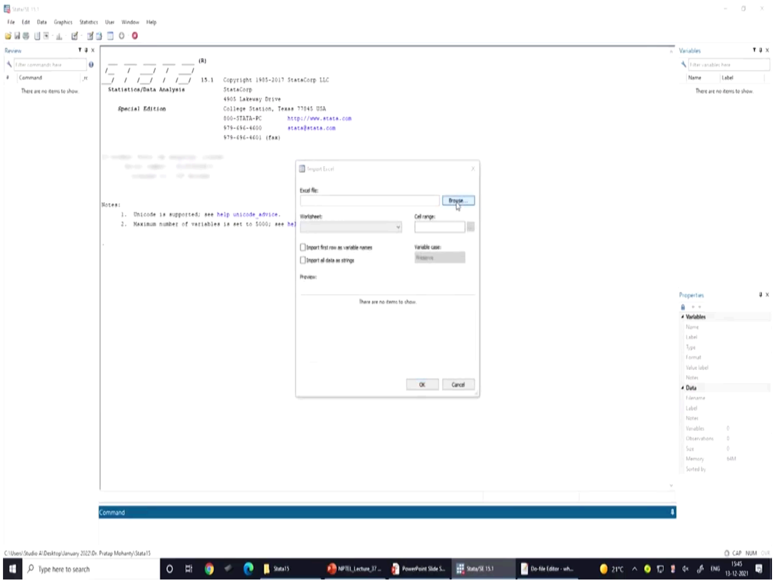
On the start of window, we are deriving results. I am going to open the first on the WHO dataset.

(Refer Slide Time: 10:02)



```
                                        (R)
  Statistics/Data Analysis      15.1   Copyright 1985-2017 StataCorp LLC
                                       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)

                        ee help unicode_advice.
                        bles is set to 5000; see help set_maxvar.
```

(Refer Slide Time: 10:05)



```
                                        (R)
  Statistics/Data Analysis      15.1   Copyright 1985-2017 StataCorp LLC
                                       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
```

Import Excel

Excel file: [ ] Browse...

Worksheet:

Cell range:

□ Import first row as variable names

□ Import all data as strings

Variable case: Preserve

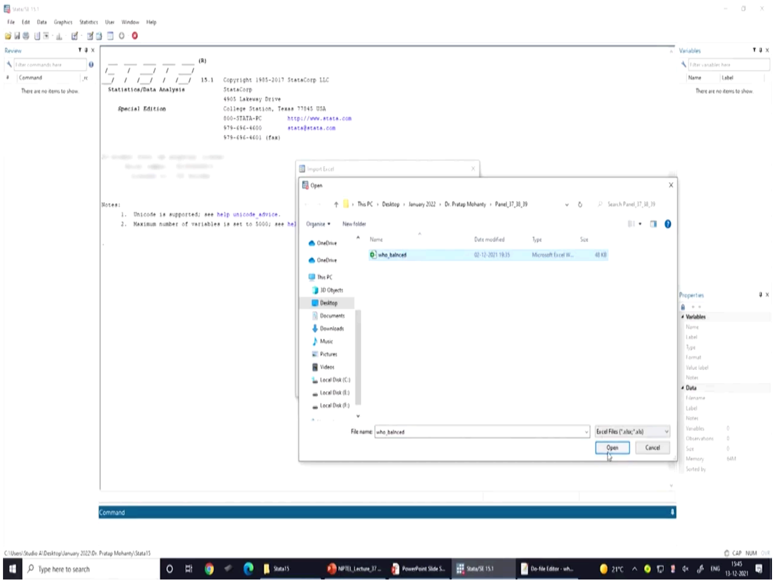
Preview:

There are no items to show.

OK Cancel

Now so I will open the file and import the data on your screen.

(Refer Slide Time: 10:07)



```
  ___  ____  ____  ____  ____ (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
```

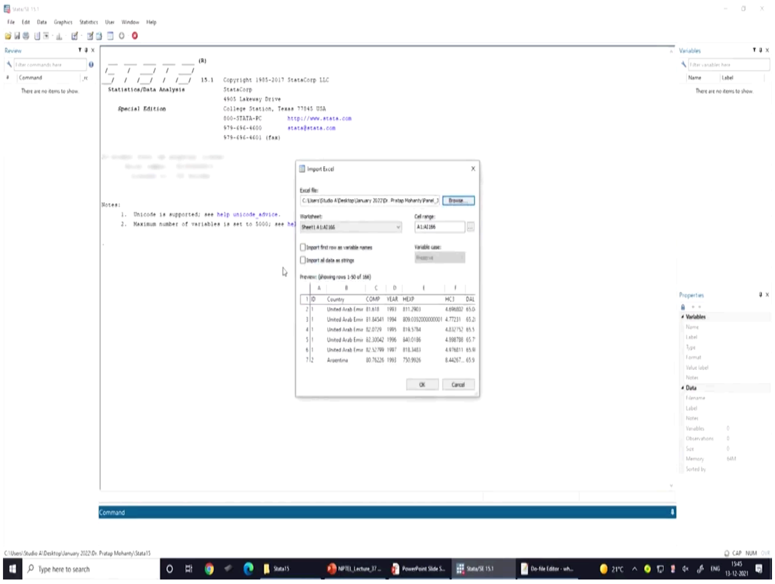
Open: This PC › Desktop › January 2022 › Dr. Pratap Mohanty › Panel_37_38_39

Name	Date modified	Type	Size
who_balnced	02-12-2021 19:35	Microsoft Excel W...	48 KB

File name: who_balnced

Excel Files (*.xlsx;*.xls)

(Refer Slide Time: 10:11)



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

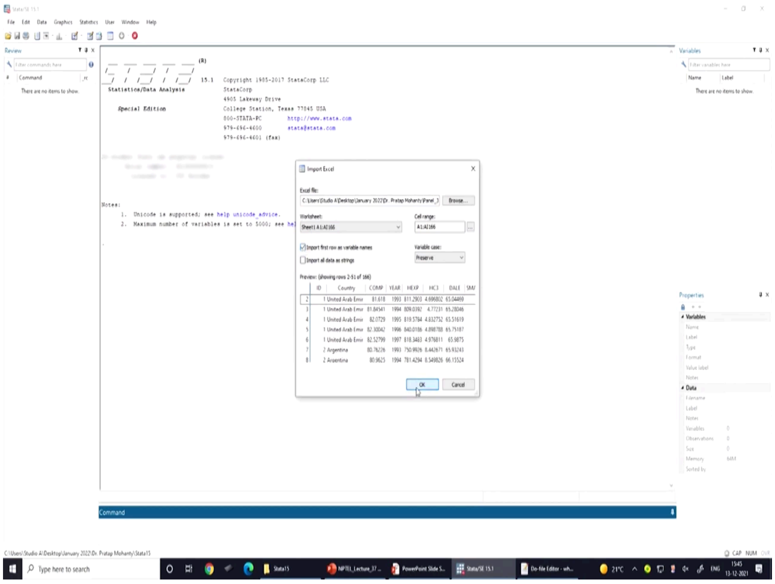
□ Import all data as strings

Preview: (showing rows 1-50 of 166)

	A	B	C	D	E	F	
1	ID	Country	COMP	YEAR	HEXP	HC3	DAL
2	1	United Arab Emir	81.618	1993	811.2903	4.696802	65.0
3	1	United Arab Emir	81.84541	1994	809.0392000000001	4.77231	65.2
4	1	United Arab Emir	82.0729	1995	819.5784	4.832752	65.5
5	1	United Arab Emir	82.30042	1996	840.0186	4.998788	65.7
6	1	United Arab Emir	82.52799	1997	818.3483	4.976811	65.9
7	2	Argentina	80.76226	1993	750.9926	8.44267...	65.9

And this is the data, we are going to show it to you or going to provide you.

(Refer Slide Time: 10:13)



```
  ___  ____  ____  ____  ____ (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
```

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

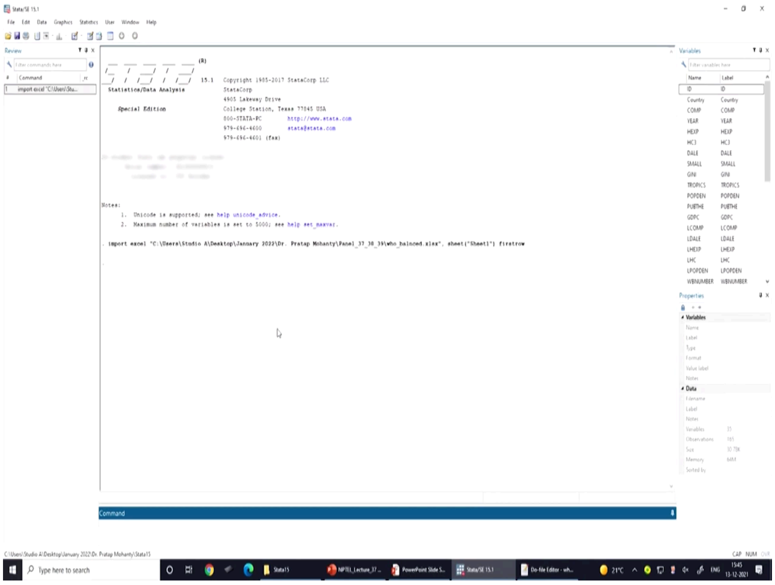
Variable case: Preserve

Preview: (showing rows 2-51 of 166)

	ID	Country	COMP	YEAR	HEXP	HC3	DALE	SMA
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	80.76226	1993	750.9926	8.442671	65.93243	
8	2	Argentina	80.9625	1994	781.4294	8.549826	66.15524	

And now we can define that which are your row variable.

(Refer Slide Time: 10:17)

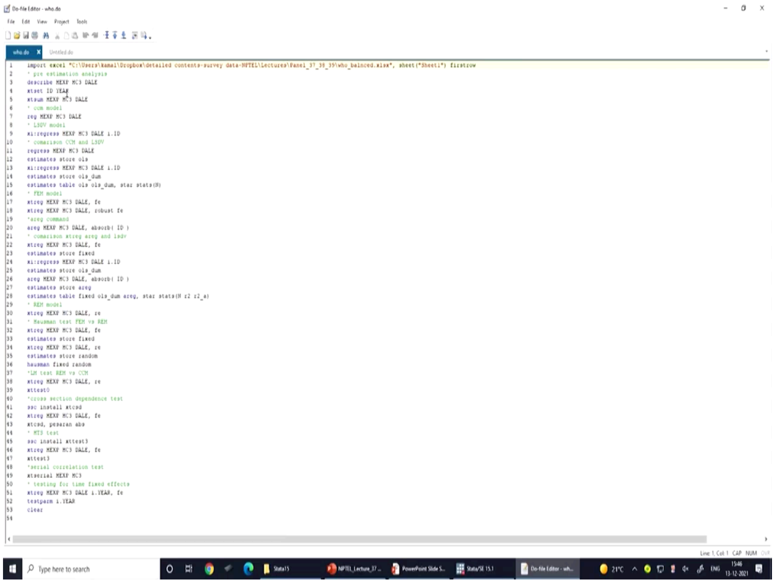


```
Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

. 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

And the row variable is basically that identifies your variable names correctly. Now we will open and share with you the constant common model analysis.

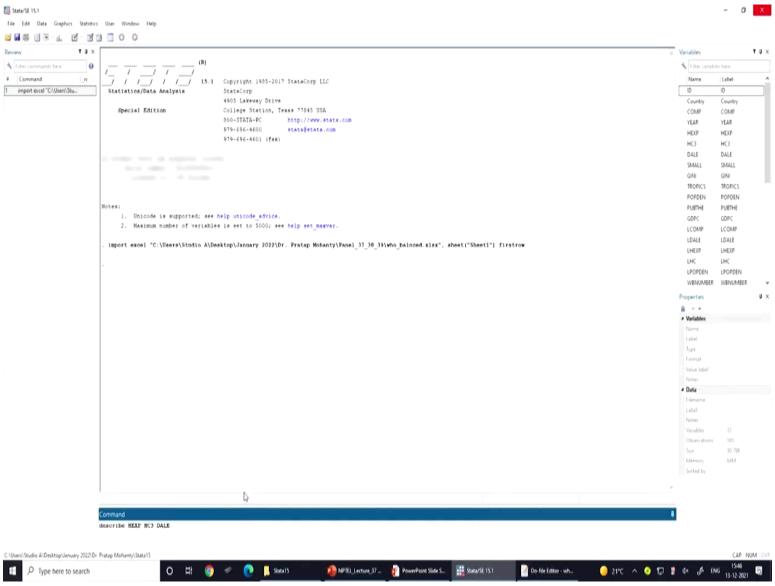
(Refer Slide Time: 10:35)



```
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 HEXP HC3 DALE
xtset ID YEAR
xtsum HEXP HC3 DALE
* ccm model
reg HEXP HC3 DALE
* LSDV model
xi:regress HEXP HC3 DALE i.ID
* comarison CCM and LSDV
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)
* FEM model
xtreg HEXP HC3 DALE, fe
xtreg HEXP HC3 DALE, robust fe
*areg command
areg HEXP HC3 DALE, absorb( ID )
* comarison xtreg areg and lsdv
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)
* REM model
xtreg HEXP HC3 DALE, re
* Hausman test FEM vs REM
xtreg HEXP HC3 DALE, fe
estimates store fixed
xtreg HEXP HC3 DALE, re
estimates store random
hausman fixed random
*LM test REM vs CCM
xtreg HEXP HC3 DALE, re
xttest0
*cross section dependence test
ssc install xtcsd
xtreg HEXP HC3 DALE, fe
xtcsd, pesaran abs
* MTS test
ssc install xttest3
xtreg HEXP HC3 DALE, fe
xttest3
*serial correlation test
xtserial HEXP HC3
* testing for time fixed effects
xtreg HEXP HC3 DALE i.YEAR, fe
testparm i.YEAR
clear
```

And in this WHO dataset we will be running some of the commands we have given here for you. And in order to cut short our time and to explain the best in the lecture we have kept a do file for your reference as well. Here the first one we have already imported and now we are doing some pre-estimation. And first pre-estimation we are doing with this data is describing the data with three important variable household expenditure is the first one that is household expenditure on health.

(Refer Slide Time: 11:09)



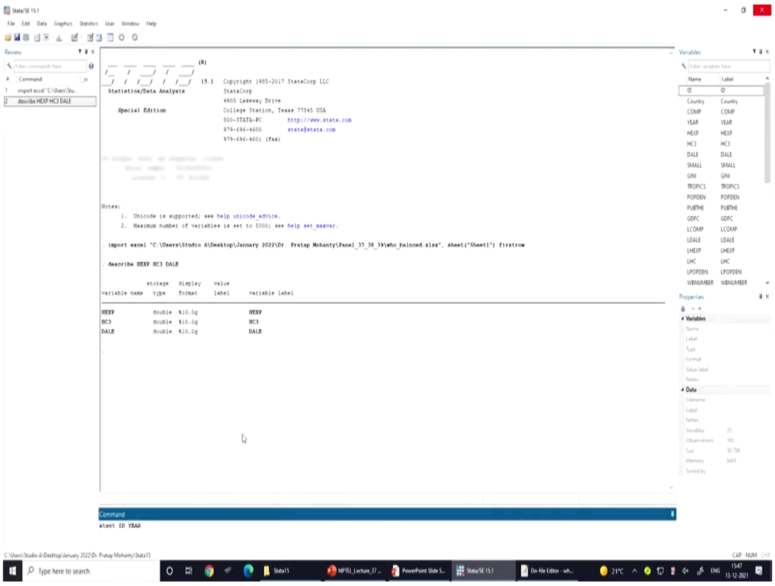
```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 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

.
```

Command

```
describe HEXP HC3 DALE
```

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

(Refer Slide Time: 11:17)



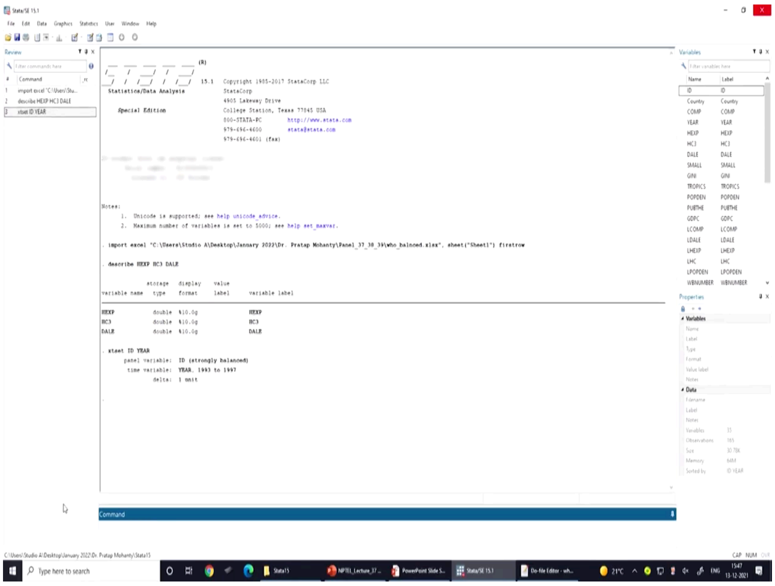
```
                                        15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis                     StataCorp
                                               4905 Lakeway Drive
     Special Edition                           College Station, Texas 77845 USA
                                               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

. describe HEXP HC3 DALE

              storage   display    value
variable name   type    format     label      variable label
---------------------------------------------------------------
HEXP            double  %10.0g                HEXP
HC3             double  %10.0g                HC3
DALE            double  %10.0g                DALE

.
```

```
xtset ID YEAR
```

So, H stands for health expenditure. So, that is here and describe is going to give you the nature of the data on your screen. Then next will be clarifying you about the extricate we need to set whether which kind of panel it is what is the panel Id's.

So, Id and the year are the two variables defined as the panel units of panel variables for us, here t is the time variable, Id is the cross sectional variable. So, now, we have defined "xt set" is mandatory for you to run at first to recognize the STATA to go with the panel data estimation.

(Refer Slide Time: 11:52)

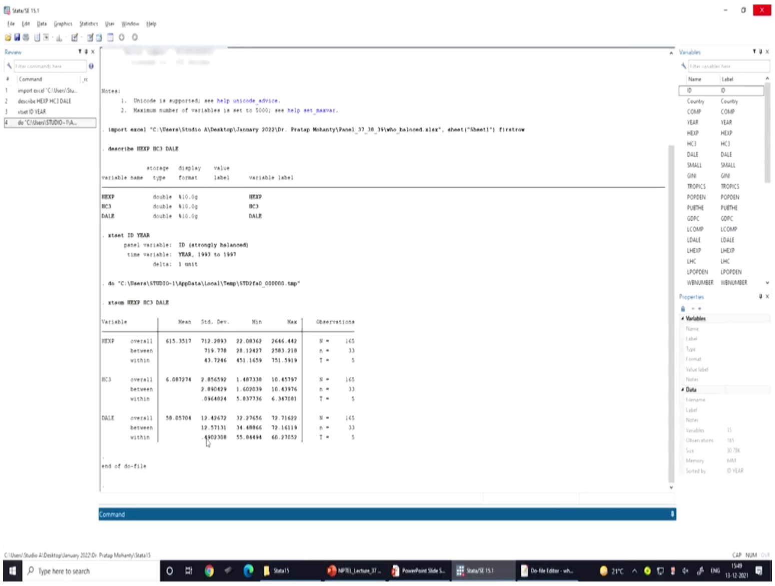


```
Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

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

. describe HEXP HC3 DALE

              storage   display    value
variable name   type    format     label      variable label
---------------------------------------------------------------
HEXP            double  %10.0g                HEXP
HC3             double  %10.0g                HC3
DALE            double  %10.0g                DALE

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

.
```

Now, this once you set it. This suggest that yes this is a strongly balanced data. So, again some of those clarification you can get it from our previous lecture as well as you can refer to my previous module. Now the data is from 93, 1993 to 1997 and the change by time is of one unit.

So, one unit time effect change is given. Now we are going to run the third important pre-estimation technique that is the summary of the data of these three important variables that is x t sum. We are going to operate xt sum here.

(Refer Slide Time: 12:42)



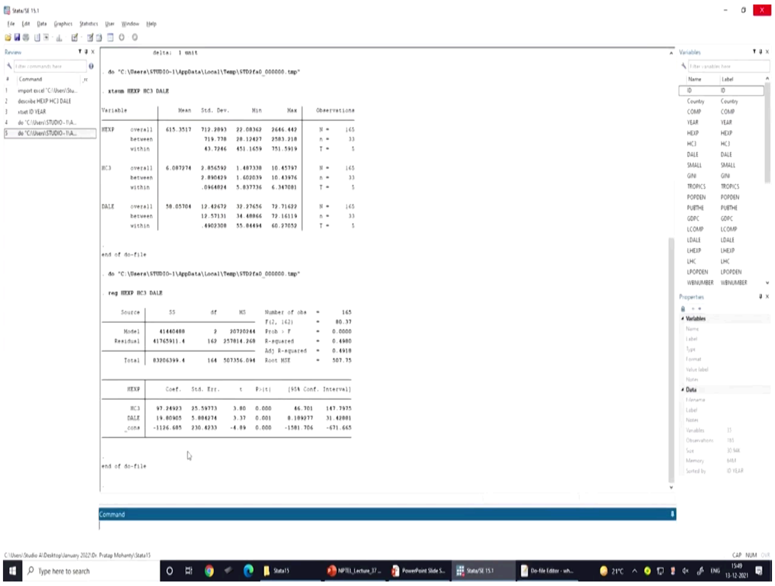
```
Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

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

. describe HEXP HC3 DALE

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
HEXP            double  %10.0g                HEXP
HC3             double  %10.0g                HC3
DALE            double  %10.0g                DALE

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

. do "C:\Users\STUDIO~1\AppData\Local\Temp\STD2fa0_000000.tmp"

. xtsum HEXP HC3 DALE

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
HEXP     overall |  615.3517   712.2093   22.00362   2646.442 |     N =     165
         between |             719.770    20.12427   2503.210 |     n =      33
         within  |              43.7246   451.1659   751.5919 |     T =       5
                 |                                            |
HC3      overall |  6.007274   2.056592   1.407330   10.45797 |     N =     165
         between |             2.090429   1.602039   10.43976 |     n =      33
         within  |             .0964024   5.037736   6.347081 |     T =       5
                 |                                            |
DALE     overall |  50.05704   12.42672   32.27656   72.71622 |     N =     165
         between |             12.57131   34.40066   72.16119 |     n =      33
         within  |             .4902300   55.04494   60.27052 |     T =       5

.
end of do-file

.
```

Now, the result is on your screen.The xt sum with these three variables that summary is giving information. So, we are getting in this panel one of the important aspect is that we are going to get the overall model within effect and between effects as well. So, since this is a pool data, it is not capturing the between and within. It is simply considering the overall estimation. Whereas, in case of fixed effect and random effect you can easily differentiate these categories.

Now, but it has given its standard deviation in each category. By mean, it has considered the overall mean, but by standard deviation there would be certainly some differences in the minimum value, maximum value, and the number of time periods are 5 in each case since it is a strongly balanced panel.

So, these are all the description for you. Now we are going to run with the proper CCM model on your screen. So, CC common constant model is a kind of pool data as I told you. So, the simple regression command is going to give you the result.

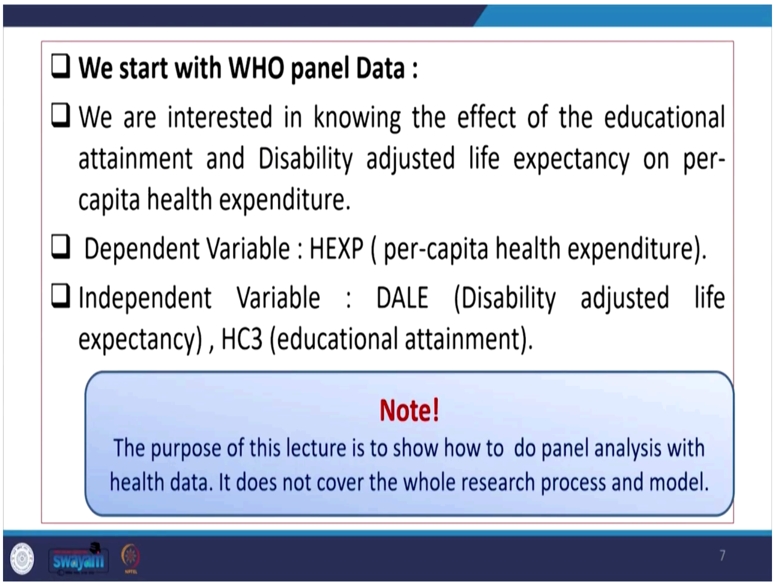
(Refer Slide Time: 13:50)



```
                  delta:  1 unit

. do "C:\Users\STUDIO-1\AppData\Local\Temp\STD2fa0_000000.tmp"

. xtsum HEXP HC3 DALE

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
HEXP     overall |  615.3517   712.2093   22.08362   2646.442 |     N =     165
         between |             719.778    28.12427   2583.218 |     n =      33
         within  |             43.7246    451.1659   751.5919 |     T =       5
                 |                                            |
HC3      overall |  6.087274   2.056592   1.487338   10.45797 |     N =     165
         between |             2.090429   1.602039   10.43976 |     n =      33
         within  |             .0964024   5.837736   6.347081 |     T =       5
                 |                                            |
DALE     overall |  58.05704   12.42672   32.27656   72.71622 |     N =     165
         between |             12.57131   34.48866   72.16119 |     n =      33
         within  |             .4902308   55.84494   60.27052 |     T =       5

.
end of do-file

. do "C:\Users\STUDIO-1\AppData\Local\Temp\STD2fa0_000000.tmp"

. reg 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 |   18.86905   5.604274     3.37   0.001     6.309277    31.42881
       _cons |  -1126.605   230.4233    -4.89   0.000    -1581.706    -671.665
------------------------------------------------------------------------------

.
end of do-file

.
```

So, on your screen we have derived the reg where the health expenditure data is the dependent variable. Now you can easily compare this result with the ordinarily square data. This is in fact, the ordinarily square model and the coefficient we can interpret easily. The P values can be interpret by the way we usually do it. So, there is no major difference as compared to the OLS model.

(Refer Slide Time: 14:28)



❑ **We start with WHO panel Data :**

❑ We are interested in knowing the effect of the educational attainment and Disability adjusted life expectancy on per-capita health expenditure.

❑ Dependent Variable : HEXP (per-capita health expenditure).

❑ Independent Variable : DALE (Disability adjusted life expectancy) , HC3 (educational attainment).

Note!

The purpose of this lecture is to show how to do panel analysis with health data. It does not cover the whole research process and model.

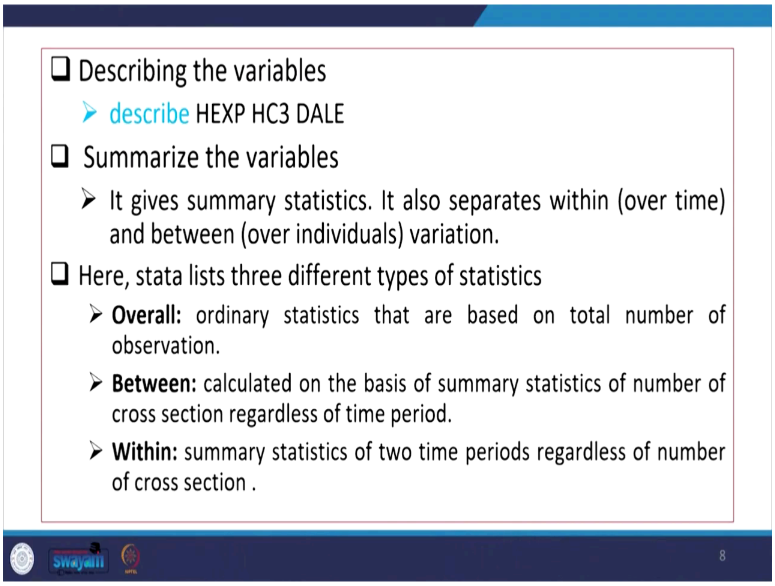
Now, let us move to our PPT once again. I will also show it with explanation in writing. So, we are explaining the WHO panel data just to observe the effect of education attainment and disability adjusted life expectancy on per capita health expenditure. So, the health expenditure variable was taken as the dependent variable. So, dependent variable is per capita health expenditure.

Whereas, the independent variable as are like educational attainment and dale disability adjusted life expectancy. The purpose of this lecture is to show how to do panel analysis with health data, it does not cover the whole research process and model.

So, we will steadily capture other directions of panel data, but at this moment we are just trying to give you the background information about a common constant model which is essentially required for discussion in order to start with fixed effect and random effect model.

Now, just to clarify I have already discussed about describe command and sum/summarize command that effects in summarizing the statistics. So, basically in summary it also separates within or over time and between variations.

(Refer Slide Time: 15:49)

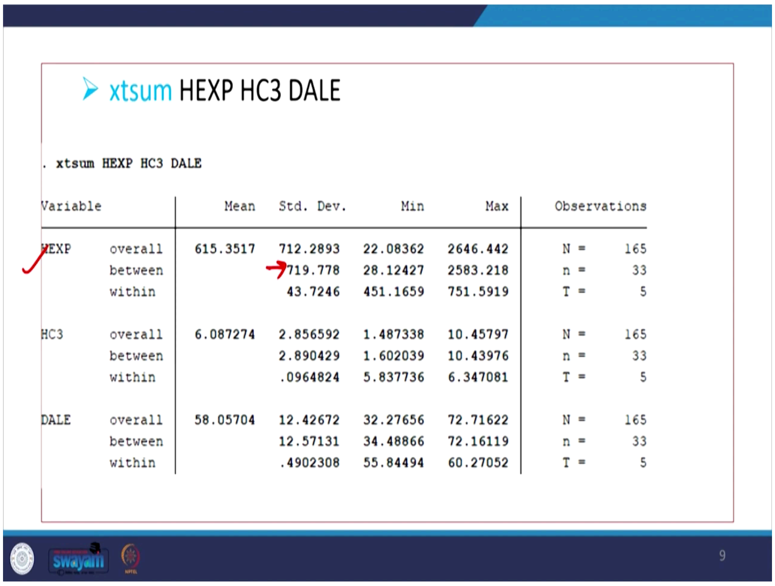


- ❑ Describing the variables
  - ➢ describe HEXP HC3 DALE
- ❑ Summarize the variables
  - ➢ It gives summary statistics. It also separates within (over time) and between (over individuals) variation.
- ❑ Here, stata lists three different types of statistics
  - ➢ **Overall:** ordinary statistics that are based on total number of observation.
  - ➢ **Between:** calculated on the basis of summary statistics of number of cross section regardless of time period.
  - ➢ **Within:** summary statistics of two time periods regardless of number of cross section .

So, the standard deviation value within and within basically over time. Because within the cross sections, how the same cross section unit is varying over time that is why it is called within. Whereas between basically between the individuals. So, how individuals are actually and their responses are varying across individuals has nothing to do with the time factor.

So, between is not taking comparison of the time whereas, within is actually emphasising the time aspects. So, these are all three types of statistics- overall, then between, and then within. Between as I told you calculated on the basis of some statistics of number of cross section regardless of the time period. And within is basically considering the time factor and casing the time changes.

(Refer Slide Time: 16:44)



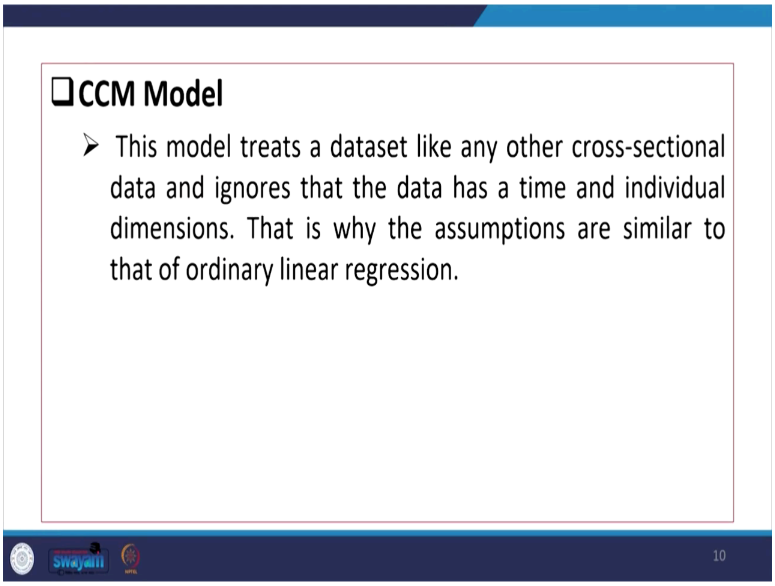
➢ xtsum HEXP HC3 DALE

```
. xtsum HEXP HC3 DALE

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
HEXP     overall |  615.3517   712.2893   22.08362   2646.442 |     N =     165
         between |             719.778    28.12427   2583.218 |     n =      33
         within  |              43.7246   451.1659   751.5919 |     T =       5
                 |                                            |
HC3      overall |  6.087274   2.856592   1.487338   10.45797 |     N =     165
         between |             2.890429   1.602039   10.43976 |     n =      33
         within  |             .0964824   5.837736   6.347081 |     T =       5
                 |                                            |
DALE     overall |  58.05704   12.42672   32.27656   72.71622 |     N =     165
         between |             12.57131   34.48866   72.16119 |     n =      33
         within  |             .4902308   55.84494   60.27052 |     T =       5
```

So, the xtsum command we have clarified already derived on your screen and you can check that in between and within effect, the variations are observed. So, some basic ideas you can take like across the individual units since it is there are so many observations 165 observations.

The between variations are more than that of the within variations. So, over the time changes are not much. So, far as per capita expenditure is concerned. Whereas, education and disabilities aspect is constant you can also check the differences accordingly.

(Refer Slide Time: 17:28)

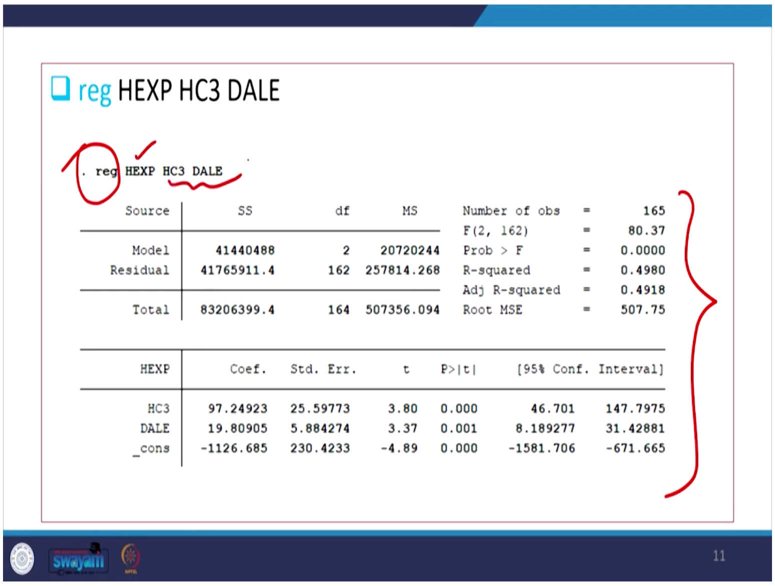


❑ CCM Model

➢ This model treats a dataset like any other cross-sectional data and ignores that the data has a time and individual dimensions. That is why the assumptions are similar to that of ordinary linear regression.

Now, the CCM model we have already derived this treats a dataset like any other cross sectional data. This ignores the data that has a time and individual dimensions that is why the assumption that are similar to that of the ordinary least square model.

(Refer Slide Time: 17:46)



❑ reg HEXP HC3 DALE

```
. reg 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
```

So, these are the results we explained and we have clarified how the coefficient could be interpreted and the P value could be compared. So, further to this you may also refer to my previous lecture as well previous lecture on handling large scale data using STATA. So, this is a similar to that of the explanation made in OLS regression that is why we carry with the command reg alright, reg dependent variable and in control variables. If any are there and accordingly you can run the regression and interpret it. There are no more things to add. Rest of the things we will explore and find out.

If you have any difficulties, we will be happy to address it. These are all for this lecture. In the next lecture we will be clarifying on fixed effect model that is going to be very interesting. And I will explain you the equation very carefully and I will also discuss about i term, t term and it will be more fascinating for you to attend.

Thank you, let me stop here.