

Stata Manual for EE425

This short manual is for EE425 students, who are completely new to Stata. Students are asked to have some level of preliminary knowledge on statistics and econometrics to understand the command usages introduced in this manual.

Stata is an easy-to-learn, fast and powerful statistical packages, which is getting popular these days. Mastering Stata will be a good weapon if you are going into research field in the near future. For more information, you can visit www.stata.com. Or other websites include

www.princeton.edu/~erp/stata/main.html

www.ats.ucla.edu/stat/stata/

This introduction is based on Stata 9 version. Thus, some commands may be incompatible in other versions. We are going to cover on :

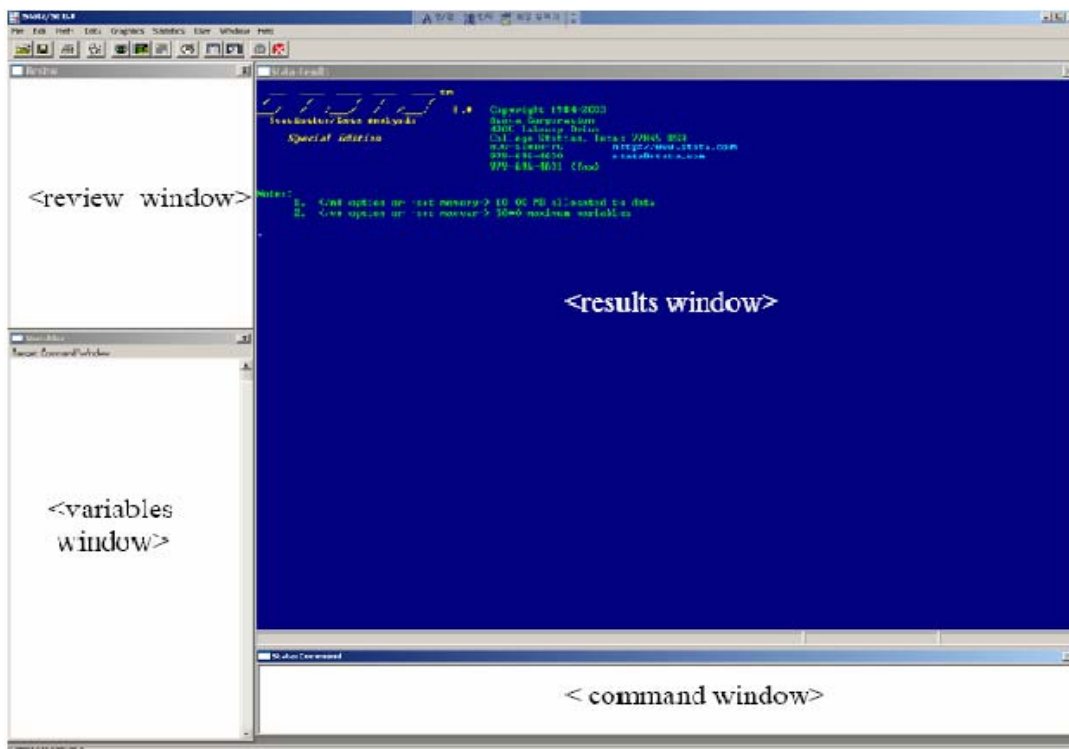
1. Preliminaries
2. Importing/Saving data into Stata
3. Data management
4. Regression
5. Example

****Data set for this subject is located at <ftp://econ.tu.ac.th/class/archan/CHAIYUTH/EE425/>**

1. Preliminaries

When you execute Stata, your screen looks like the one below, which is my default windowing. Basically, you can work with four windows:

- review window: shows your command history. By clicking the line, you don't have to retype lengthy and similar commands on command window.
- results window: shows the output from your commands
- command window: type your command here and enter to execute
- variables window: shows the variables in the dataset you're currently working with



You may open/close these windows with resizing, and save it as a default windowing preference in Prefs menu.

- **help** command: Naturally, the most helpful command in Stata is **help**. The usage is just to type **help [command name]** in command window and enter. For example, if you'd like to know about regression, go help → search... and look for 'regression'. Then it shows many commands related with regression. Among them, you can find a command **regress** (shortcut **reg**). If you need more detail, just type **help regress**, then results window shows the usage for regress command.

- **log** command: when you need to record the output with your work history, use this command. For example,

log using ps1 : it will make a log file named 'ps1' in your default working directory, which records the commands and output

log close : it will close ps1 log file

log using ps1, append: option append is used when you append more results to the end of an already existing log file

For more information, type **help log**.

2. Importing/saving data into Stata

You can import data into Stata in the forms of Stata dataset (with extension name 'dta'), text/ascii file, or csv file from Excel.

- **use** command: when you have a file in Stata dataset format, named ps1.dta in c:\work:

use c:\work\ps1.dta

- **infile** command: your data file should not have variable names on the first line. You can name them if you want when you import data. Or you can use **rename [old name] [new name]** after reading data:

infile using c:\work\ps1.raw or

infile wage sex education region industry using c:\work\ps1.raw

- **insheet** command: this command is very useful when you read data from a spreadsheet. In many cases, you will have data in the format of Excel spreadsheet. To read it into Stata, you need to save xls file as csv file in save as menu and then use

insheet using command:

insheet using c:\work\ps1.csv

- **save** command: you can save the current file in memory by using this command:

```
save ps1_1
```

It will save with the name of ps1_1.dta in your working directory

3. Data management

- **summarize** (or **sum**) command: it gives summary statistics of the variables. And **detail** options gives more information on distribution:

```
sum
```

```
sum wage
```

```
sum wage education, detail
```

- **describe** (or **des**) command: it gives contents of the data in memory, such as on the types (string or numeric etc) and length:

```
des
```

```
des wage
```

```
des wage education
```

- **browse** command: it shows your data in a spreadsheet format. Or just click 'Data Browser' icon below window menu:

The screenshot shows the Stata Data Browser window. The main area displays a spreadsheet of data with columns labeled from _id to _b100. The data is organized into rows, with the first row starting with _id: 1 and ending with _b100: 2.1973. The left sidebar shows a list of variables including _id, _weight, _wtsqr, _b1, _b2, _b3, _b4, _b5, _b6, _b7, _b8, _b9, _b10, _b11, _b12, _b13, _b14, _b15, _b16, _b17, _b18, _b19, _b20, _b21, _b22, _b23, _b24, _b25, _b26, _b27, _b28, _b29, _b30, _b31, _b32, _b33, _b34, _b35, _b36, _b37, _b38, _b39, _b40, _b41, _b42, _b43, _b44, _b45, _b46, _b47, _b48, _b49, _b50, _b51, _b52, _b53, _b54, _b55, _b56, _b57, _b58, _b59, _b60, _b61, _b62, _b63, _b64, _b65, _b66, _b67, _b68, _b69, _b70, _b71, _b72, _b73, _b74, _b75, _b76, _b77, _b78, _b79, _b80, _b81, _b82, _b83, _b84, _b85, _b86, _b87, _b88, _b89, _b90, _b91, _b92, _b93, _b94, _b95, _b96, _b97, _b98, _b99, _b100. The top toolbar includes icons for File, Edit, View, and other standard spreadsheet functions.

- **generate** (or **gen**) command: You can make new variables with **generate**:

```

gen education2 = education*education
gen wage_month = wage/12
gen id = _n
gen wage_lag = wage[_n-1]

```

4. Regression

- **regress** (or **reg**) command: As you expect, most important command in Econ 210 class! It fits linear models. The usage is **reg** [*dependent variable*] [*explanatory variables*]:

```

reg wage education
reg wage education sex region
reg wage education if wage > 50000
reg wage education if region == 1

```

- **predict** command: You can generate fitted values or residuals using this command.

```

predict [variable name] will create predicted value from fitted regression model and
predict [variable name], residual will make residual values:
predict yhat
predict e, residual or predict e, resi

```

- **test** command: This is for testing your hypothesis. Option **accumulate** allows a hypothesis to be tested jointly with the previously tested hypotheses:

```

test _cons = 0 : test if the intercept is statistically equal to zero
test education = 0.5
test _cons = 0
test education = 0.5, accumulate : joint test

```

scatter diagram and fitted line graph

Before doing the OLS, if you would like to see how the data looks by using this command **Twoway scatter** [*variable on Y axis*][*variable on X axis*]

In the case if you want to put the name to the diagram, you can put comma and add title then the name at the end of this command. **Twoway scatter** [*variable on Y axis*][*variable on X axis*],title(name)

```

Twoway scatter y x
Twoway scatter y x ,title(relationship between y and x)

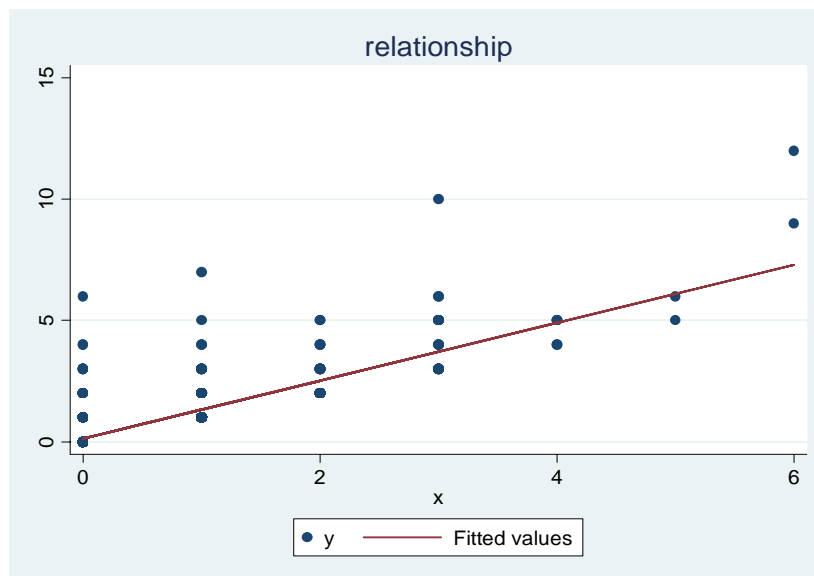
```

If you want to see the fitted line, you can use **twoway line [fitted value of dependent variable][independent variable]**

Twoway line yhat x if yhat is fitted value of y

If you want to put scatter diagram and straight line(fitted line) together, you can use **Twoway (scatter [variable on y axis][variable on x axis])(line [fitted value][independent variable], title(name))**

Twoway (scatter y x) (line yhat x), title(relationship between y and x)



- **Confidence interval**

Command : **set level ()** ; use to set level of significance or confidence interval level
ex. set level 99 means you set 99% confidence interval.

set level 99

Set level 90

Command : **ci () () ()....** ; use to construct confidence interval of variables with respect to level you set ex. ci yhat wage age means you construct confidence interval of predicted Y, wage and age.

ci wage age

ci yhat

ci x y z

Note: Before you construct confidence interval, you should set level of significance first.