

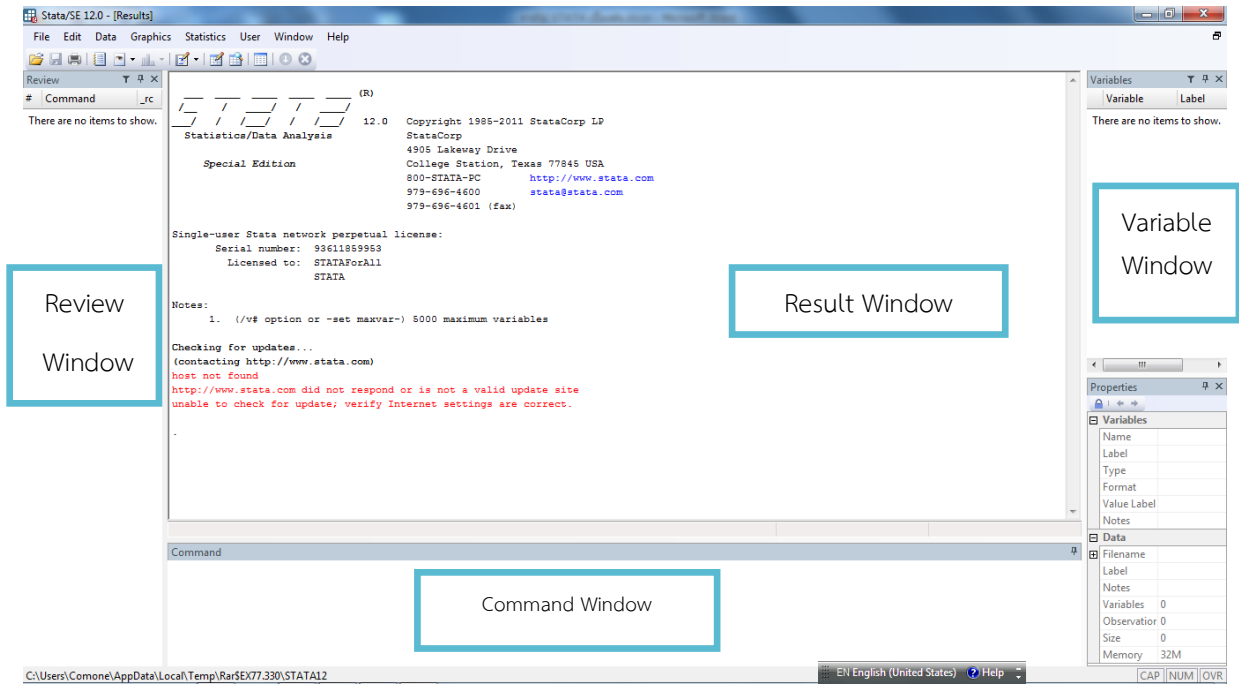
STATA Usage basic 1

- STATA Window
- Saving data in Log File
- Using the data – Importing or creating data
- Basic commands and the explanations
 - Display
 - Generate
 - Graphing
 - Scatter Plot
 - Summarize
 - Regression
 - Predict Yhat Uhat

NOTE - use only lower case letter in STATA

- Can use abbreviation for some command
- This program does not read Thai.
- Notation usage: more than or equal to \geq , less than or equal to \leq , equal to $==$
- Deleting variables: Open data editor -> selecting the variable you want to delete -> right click -> Drop selected data

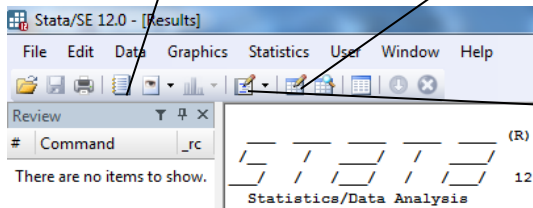
- STATA Window




Log begin: to start saving all of the results from your Stata session

Data Editor: is the menu for creating data

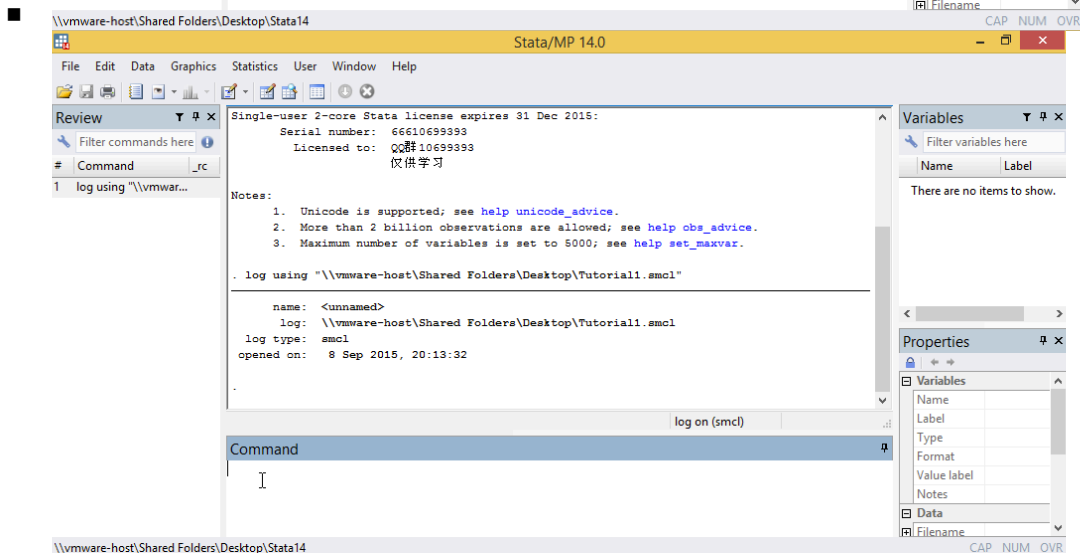
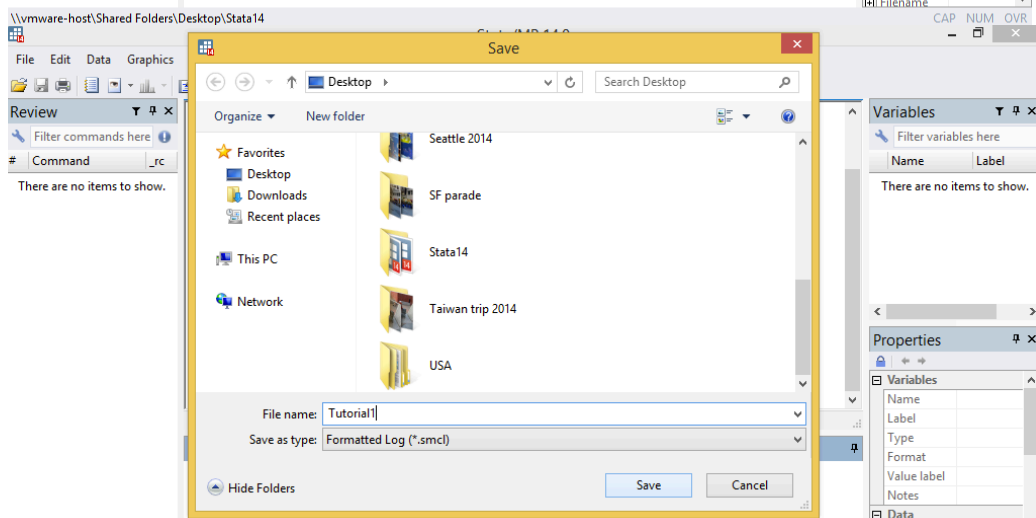
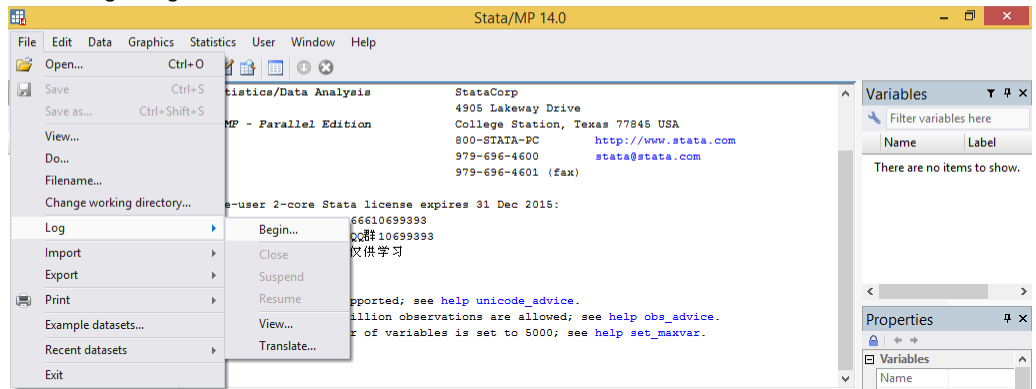
Do-file:



- Saving data in Log File: can be done by 2 methods (*for both methods choose .log)

- Quick Menu (Log Begin )

- File>Log>Begin

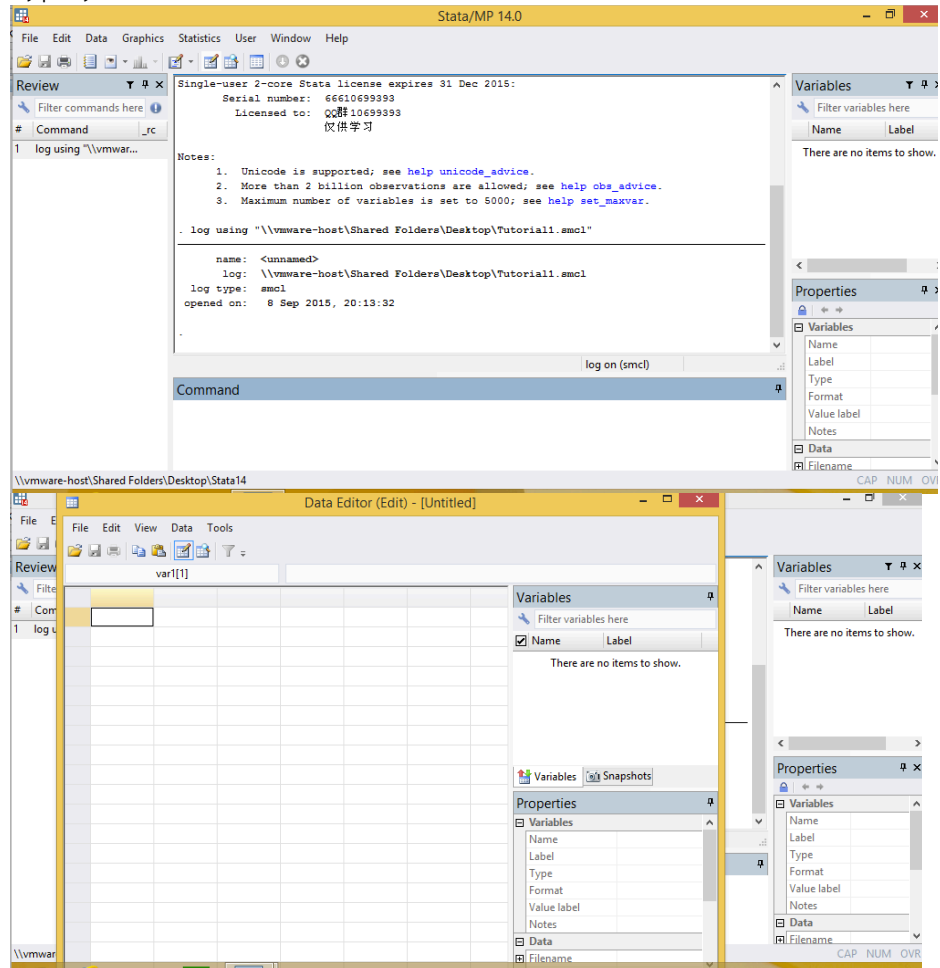


- Warning: when finish your session do not forget to Close (File>Log>Close)

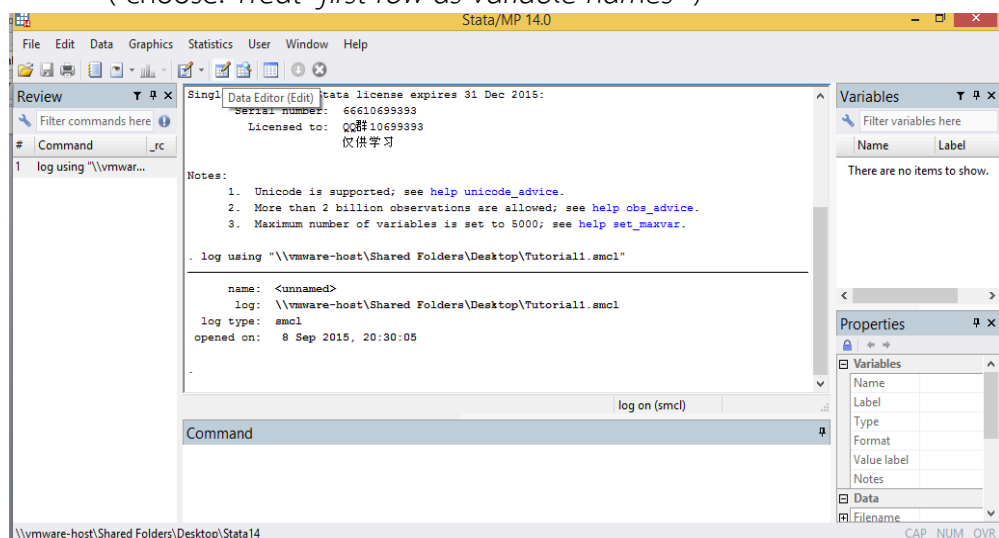
- Using the data – Importing or creating data

- Data Editor  is used for opening table for creating data.

- Type your data into the table



- Add data by Copy-Paste from Excel
(choose: *Treat first row as variable names*)



Tutorial1 data - Microsoft Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW ACROBAT Sign in

Clipboard Font Alignment Number Styles Cells

A1 : X ✓ fx year

	A	B	C	D	E	F	G	H
1	year	consumption	gdp					
2	1990	1.888117	3.37347					
3	1991	1.98795	3.656936					
4	1992	2.184372	3.994462					
5	1993	2.35953	4.340992					
6	1994	2.545363	4.688137					
7	1995	2.756101	5.068827					
8	1996	2.894826	5.355335					
9	1997	2.854002	5.20787					
10	1998	2.562901	4.810316					
11	1999	2.667109	5.030258					
12	2000	2.854801	5.254391					
13	2001	3.023108	5.435365					
14	2002	3.211201	5.769578					
15	2003	3.447105	6.184371					

Sheet1

READY AVERAGE: 668 COUNT: 45 SUM: 28,056 100%

Data Editor (Edit) - [Untitled]

File Edit View Data Tools

var[1]

- Copy Ctrl+C
- Paste Ctrl+V
- Paste special... Ctrl+Alt+V
- Select all Ctrl+A
- Data
- Hide selected variables
- Show only selected variables
- Show entire dataset
- Preferences...
- Font...

Variables

Filter variables here

Name Label

There are no items to show.

Properties

Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

Filename

Data Editor (Edit) - [Untitled]

File Edit View Data Tools

var[1]

Stata

The first row on the Clipboard contains values that can be used as valid variable names.

Do you want to treat the first row as variable names or data?

Variable names Data Cancel

Variables

Filter variables here

Name Label

There are no items to show.

Properties

Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

Filename

Data Editor (Edit) - [Untitled]

	year[1]	1990		
	year	consumption	gdp	
1	1990	1.888117	3.37347	
2	1991	1.98795	3.656936	
3	1992	2.184372	3.994462	
4	1993	2.35953	4.340992	
5	1994	2.545363	4.688137	
6	1995	2.756101	5.068827	
7	1996	2.894826	5.355335	
8	1997	2.854002	5.20787	
9	1998	2.562901	4.810316	
10	1999	2.667109	5.030258	
11	2000	2.854801	5.254391	
12	2001	3.023108	5.435365	
13	2002	3.211201	5.769578	
14	2003	3.447105	6.184371	
15	2004	3.703164	6.573325	
16	2005x	3.858725	6.848605	
17	2006x	3.9668	7.188839	
18	2007x	4.014069	7.579559	

Variables

- Name Label
- year year
- consumption consumption
- gdp gdp

Properties

Variables

Name	year
Label	year
Type	str16
Format	%16s
Value label	
Notes	

- Data>Data Editor>Data Editor(Edit)
- Import File from Excel
 - File>Import>Excel Spreadsheet (choose: *Import first row as variable*)

Stata/MP 14.0

File Edit Data Graphics Statistics User Window Help

Import

- Excel spreadsheet (*.xls;*.xlsx)
- Text data (delimited, *.csv, ...)
- Text data in fixed format
- Text data in fixed format with a dictionary
- Unformatted text data
- SAS XPORT
- Haver Analytics database
- ODBC data source
- XML data

Command

Stata/MP 14.0

Import Excel

Excel file: Browse...

Worksheet: Cell range:

Import first row as variable names

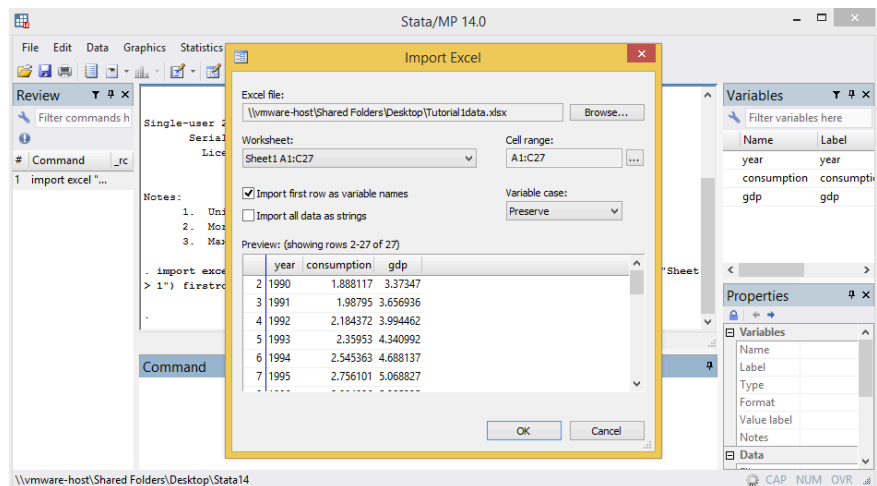
Import all data as strings

Variable case: Preserve

Preview

There are no items to show.

OK Cancel

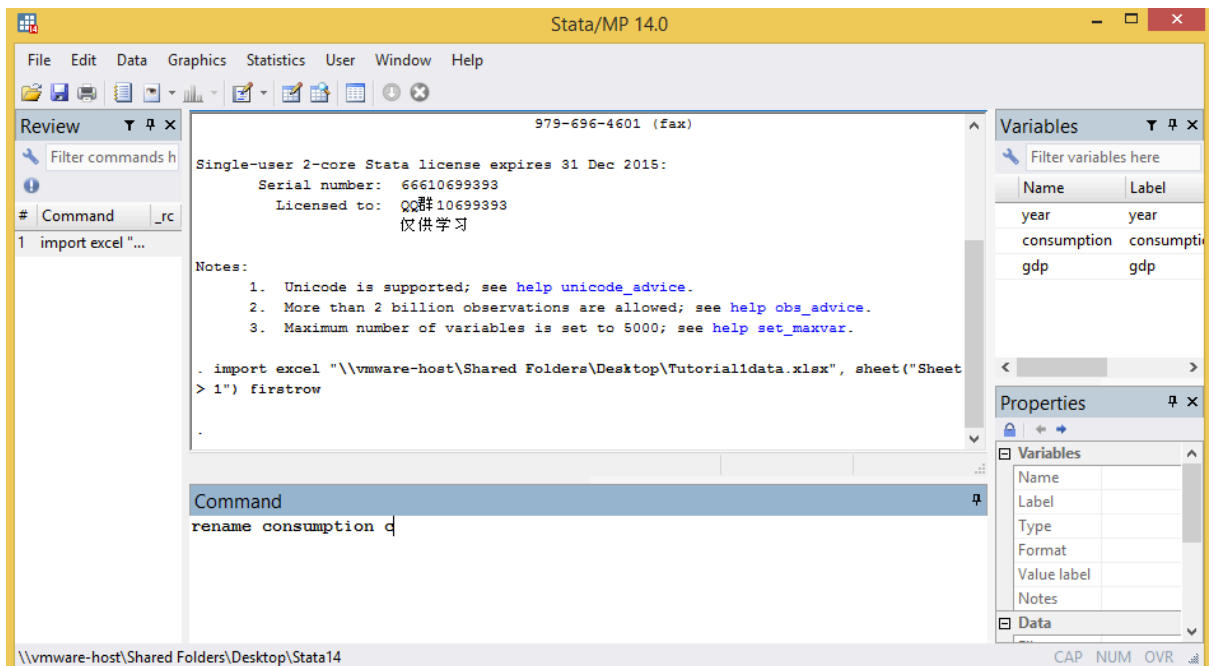


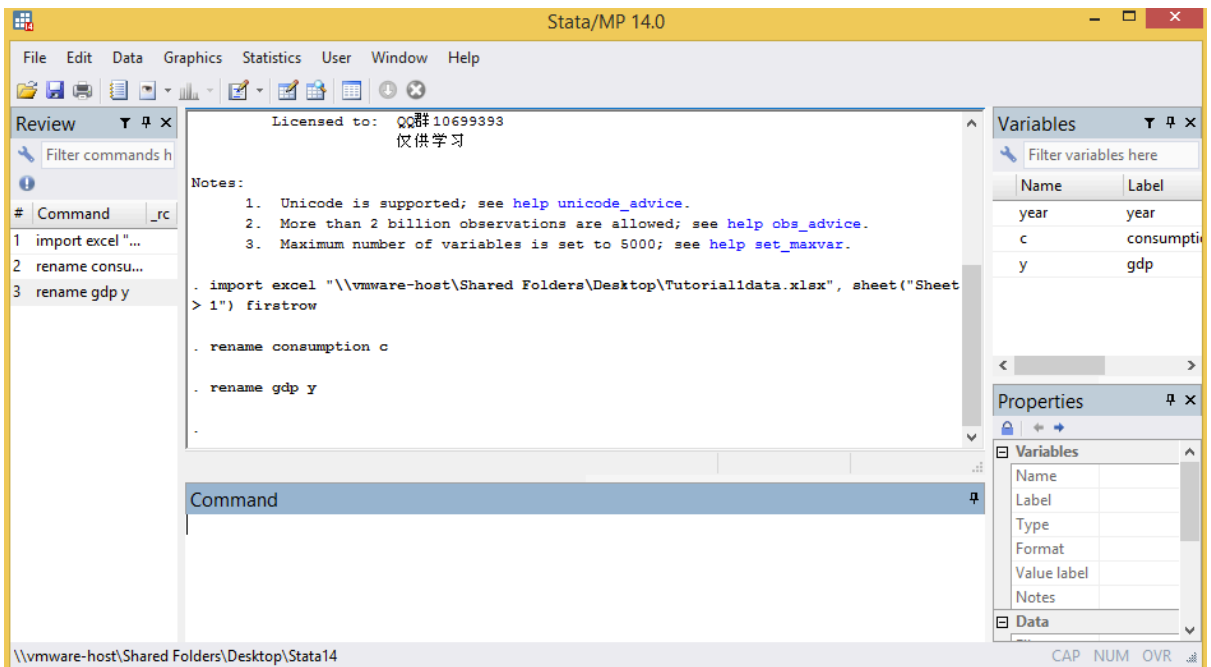
- Basic commands and the explanations

- **Display**: is the basic command used for mathematical calculation, for example: “display 5*5” will result in 25

```
. display 5*5
25
```

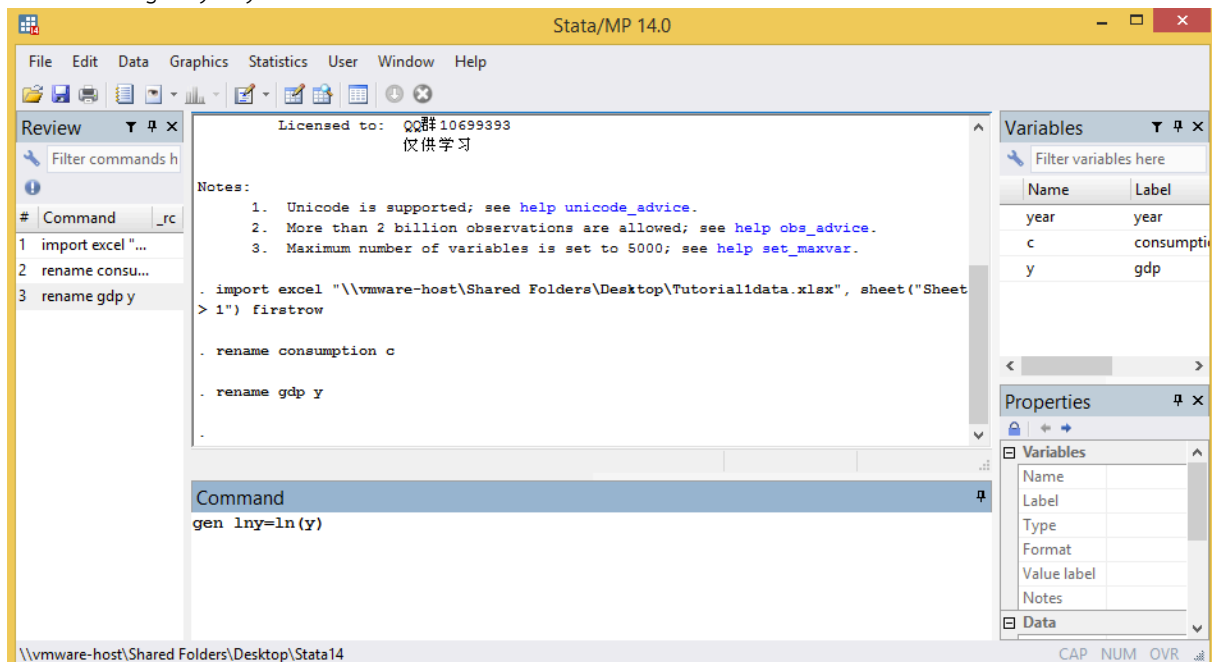
- **Rename** : changing the variable names (rename original name new name)





(or open Data Editor>Click the variable you want to rename
 Toolbar on the right > click name> and type the new name
 then press Enter)

- **Generate:** for generating variables, eg.
 - gen lny=ln(y)
 - gen yc=y*c



The screenshot shows a data editor window with a dataset containing 18 rows of data. The columns are labeled 'year', 'c', 'y', 'lny', and 'yc'. The 'year' column contains years from 1990 to 2007. The 'c' column represents consumption, 'y' represents GDP, 'lny' represents the natural log of GDP, and 'yc' represents the natural log of consumption. The right sidebar shows the 'Variables' panel with a list of variables and their labels, and the 'Properties' panel showing the properties of the selected variables.

	year	c	y	lny	yc
1	1990	1.888117	3.37347	1.215942	6.369506
2	1991	1.98795	3.656936	1.296626	7.269806
3	1992	2.184372	3.994462	1.384909	8.725391
4	1993	2.35953	4.340992	1.468103	10.2427
5	1994	2.545363	4.688137	1.545035	11.93301
6	1995	2.756101	5.068827	1.623109	13.9702
7	1996	2.894826	5.355335	1.678093	15.50276
8	1997	2.854002	5.20787	1.650171	14.86327
9	1998	2.562901	4.810316	1.570763	12.32836
10	1999	2.667109	5.030258	1.615471	13.41625
11	2000	2.854801	5.254391	1.659064	15.00024
12	2001	3.023108	5.435365	1.692927	16.4317
13	2002	3.211201	5.769578	1.752599	18.52728
14	2003	3.447105	6.184371	1.822025	21.31818
15	2004	3.703164	6.573325	1.88302	24.3421
16	2005	3.858725	6.848605	1.924045	26.42688
17	2006	3.9668	7.188839	1.97253	28.51669
18	2007	4.014069	7.579559	2.025455	30.42487

- **Label:** explanations for the variables (label variable variable's name "explanation")

```
. label var y "GDP (in trillion baht)"
```

```
. label var c "consumption (in trillion baht)"
```

(when be use Label command, the explanation will show in here)

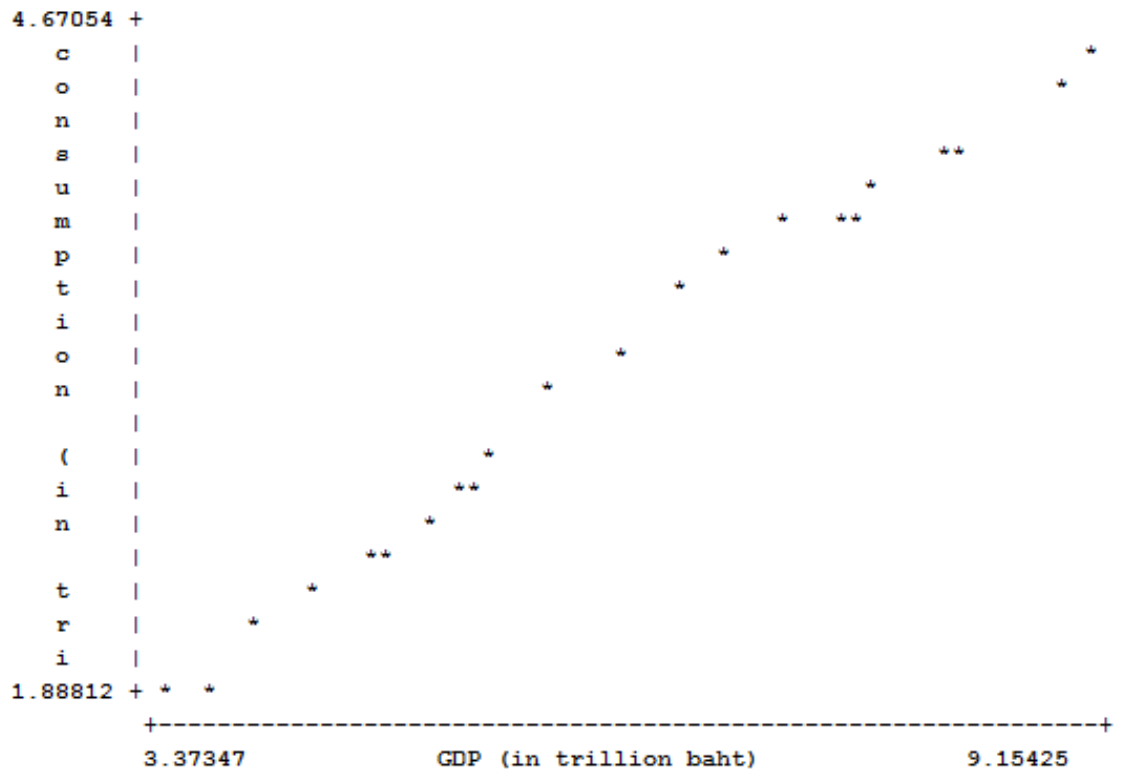
The screenshot shows the 'Variables' panel in a software interface. It contains a search bar and a table with two columns: 'Name' and 'Label'. The variables listed are 'year', 'c', 'y', 'lny', and 'yc'. The labels for 'c' and 'y' are 'consumption (in trillion baht)' and 'GDP (in trillion baht)' respectively.

Name	Label
year	year
c	consumption (in trillion baht)
y	GDP (in trillion baht)
lny	
yc	

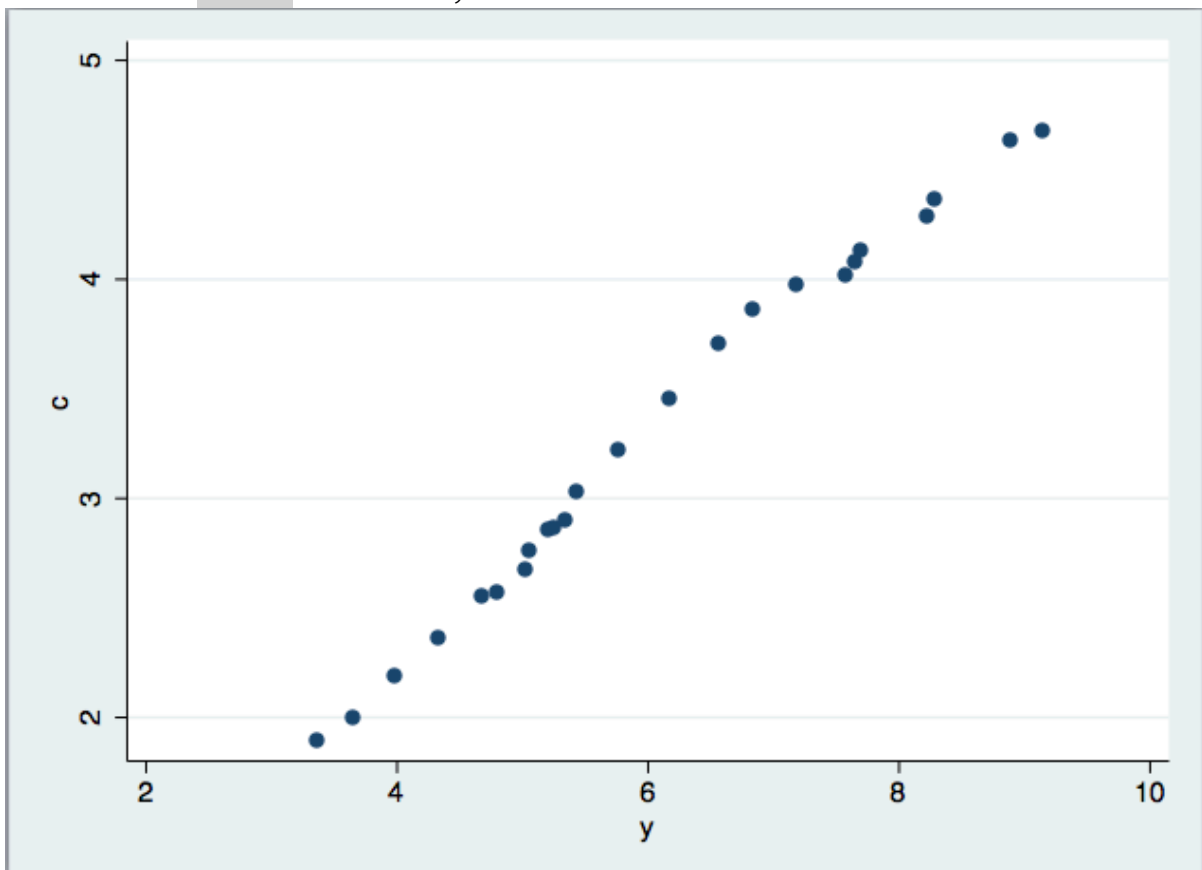
■ Graphing:

Plot : (plot c y)

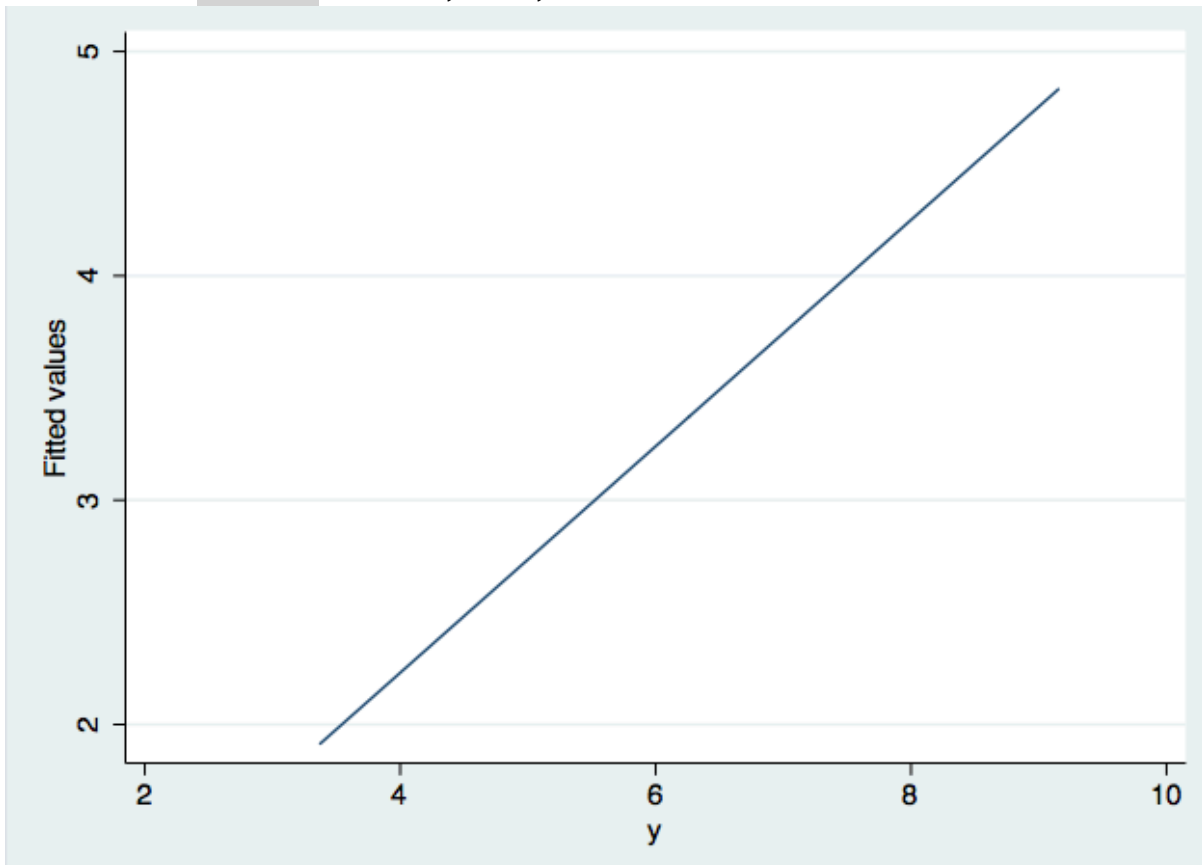
. plot c y



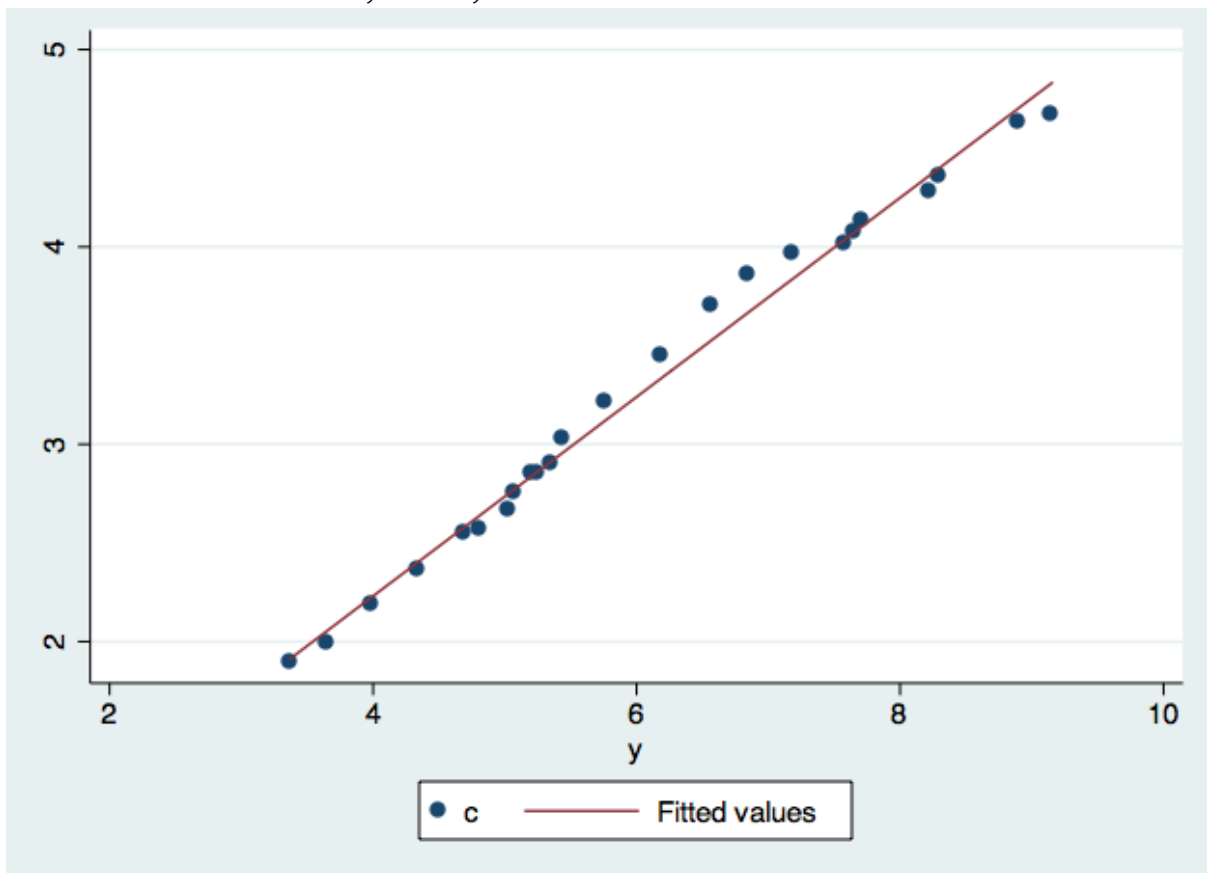
Scatter : (scatter c y)



Linear fit : (twoway lfit c y)



tw (scatter c y) (lfit c y)



- **Summarize:** gives basic statistics (Mean, Std deviation, Max, Min for each of the variables)

```
. sum y c
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	24	6.096136	1.682742	3.37347	9.154247
c	24	3.288345	.8536153	1.888117	4.670544

(or: Statistic > Summarize > Summary > Summary statistics > choose the variable (can choose multiple at a time) > ok)

- **Regression:** (reg c y)

Keynesian consumption model:

$$C_t = \alpha + \beta Y_t + u_t$$

where C is consumption expenditure

Y is income

. reg c y

Source	SS	df	MS	Number of obs	=	24
Model	16.5855074	1	16.5855074	F(1, 22)	=	2101.23
Residual	.173650864	22	.007893221	Prob > F	=	0.0000
Total	16.7591583	23	.728659057	R-squared	=	0.9896
				Adj R-squared	=	0.9892
				Root MSE	=	.08884

c	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
y	.5046414	.0110089	45.84	0.000	.4818103 .5274726
_cons	.2119818	.0695191	3.05	0.006	.067808 .3561556

$$\hat{C}_t = 0.2120 + 0.5046 Y_t$$

Interpreting the results:

1. When Y increase by 1 unit, C increase by 0.5046 unit.
2. When Y is equal to zero, C is 0.2120 unit.

Hypothesis testing:

- Using the give t values and p-value to determine the significance of the explanatory variables
- Both Y and the constant terms are statistically significant.

Confidence interval

-calculation: $\hat{b} \pm t_{\frac{\alpha}{2}} se(\hat{b})$

- For the interpretation of confidence interval, for example, the confidence level of 95 percent means, 95 out of 100 cases will include the true parameter.

R-square: variation in Y can explain variation in C by 98.96%

(Statistic>Linear Models >Linear Regression> choose dependent and independent variables > ok)

- Predict **Yhat** : (predict chat)

```
. predict chat
(option xb assumed; fitted values)
```

Data Editor (Browse) - [Untitled]

File Edit View Data Tools

1C 1,9143745

	year	c	y	lny	yc	chat
1	1990	1.888117	3.37347	1.215942	6.369506	1.914374
2	1991	1.98795	3.656936	1.296626	7.269806	2.057423
3	1992	2.184372	3.994462	1.384909	8.725391	2.227753
4	1993	2.35953	4.340992	1.468103	10.2427	2.402626
5	1994	2.545363	4.688137	1.545035	11.93301	2.57781
6	1995	2.756101	5.068827	1.623109	13.9702	2.769922
7	1996	2.894826	5.355335	1.678093	15.50276	2.914506
8	1997	2.854002	5.20787	1.650171	14.86327	2.840089
9	1998	2.562901	4.810316	1.570763	12.32836	2.639467
10	1999	2.667109	5.030258	1.615471	13.41625	2.750458
11	2000	2.854801	5.254391	1.659064	15.00024	2.863565
12	2001	3.023108	5.435365	1.692927	16.4317	2.954892
13	2002	3.211201	5.769578	1.752599	18.52728	3.12355
14	2003	3.447105	6.184371	1.822025	21.31818	3.332872
15	2004	3.703164	6.573325	1.88302	24.3421	3.529154
16	2005x	3.858725	6.848605	1.924045	26.42688	3.668072
17	2006x	3.9668	7.188839	1.97253	28.51669	3.839768
18	2007x	4.014069	7.579559	2.025455	30.42487	4.036941

Variables

Filter variables here

<input checked="" type="checkbox"/>	Name	Label
<input checked="" type="checkbox"/>	year	year
<input checked="" type="checkbox"/>	c	consumption (in t...
<input checked="" type="checkbox"/>	y	GDP (in trillion ba...
<input checked="" type="checkbox"/>	lny	
<input checked="" type="checkbox"/>	yc	
<input checked="" type="checkbox"/>	chat	Fitted values

Properties

Variables	
Name	chat
Label	Fitted values
Type	float
Format	%9.0g
Value label	
Notes	

Uhat : (predict uhat, residual)

. predict uhat, residual

Data Editor (Browse) - [Untitled]

File Edit View Data Tools

1C 1,9143745

	year	c	y	lny	yc	chat
1	1990	1.888117	3.37347	1.215942	6.369506	1.914374
2	1991	1.98795	3.656936	1.296626	7.269806	2.057423
3	1992	2.184372	3.994462	1.384909	8.725391	2.227753
4	1993	2.35953	4.340992	1.468103	10.2427	2.402626
5	1994	2.545363	4.688137	1.545035	11.93301	2.57781
6	1995	2.756101	5.068827	1.623109	13.9702	2.769922
7	1996	2.894826	5.355335	1.678093	15.50276	2.914506
8	1997	2.854002	5.20787	1.650171	14.86327	2.840089
9	1998	2.562901	4.810316	1.570763	12.32836	2.639467
10	1999	2.667109	5.030258	1.615471	13.41625	2.750458
11	2000	2.854801	5.254391	1.659064	15.00024	2.863565
12	2001	3.023108	5.435365	1.692927	16.4317	2.954892
13	2002	3.211201	5.769578	1.752599	18.52728	3.12355
14	2003	3.447105	6.184371	1.822025	21.31818	3.332872
15	2004	3.703164	6.573325	1.88302	24.3421	3.529154
16	2005x	3.858725	6.848605	1.924045	26.42688	3.668072
17	2006x	3.9668	7.188839	1.97253	28.51669	3.839768
18	2007x	4.014069	7.579559	2.025455	30.42487	4.036941

Variables

Filter variables here

<input checked="" type="checkbox"/>	Name	Label
<input checked="" type="checkbox"/>	year	year
<input checked="" type="checkbox"/>	c	consumption (in t...
<input checked="" type="checkbox"/>	y	GDP (in trillion ba...
<input checked="" type="checkbox"/>	lny	
<input checked="" type="checkbox"/>	yc	
<input checked="" type="checkbox"/>	chat	Fitted values

Properties

Variables	
Name	chat
Label	Fitted values
Type	float
Format	%9.0g
Value label	
Notes	