

Computer Lab in Economics Master in International Economics Introduction to Stata

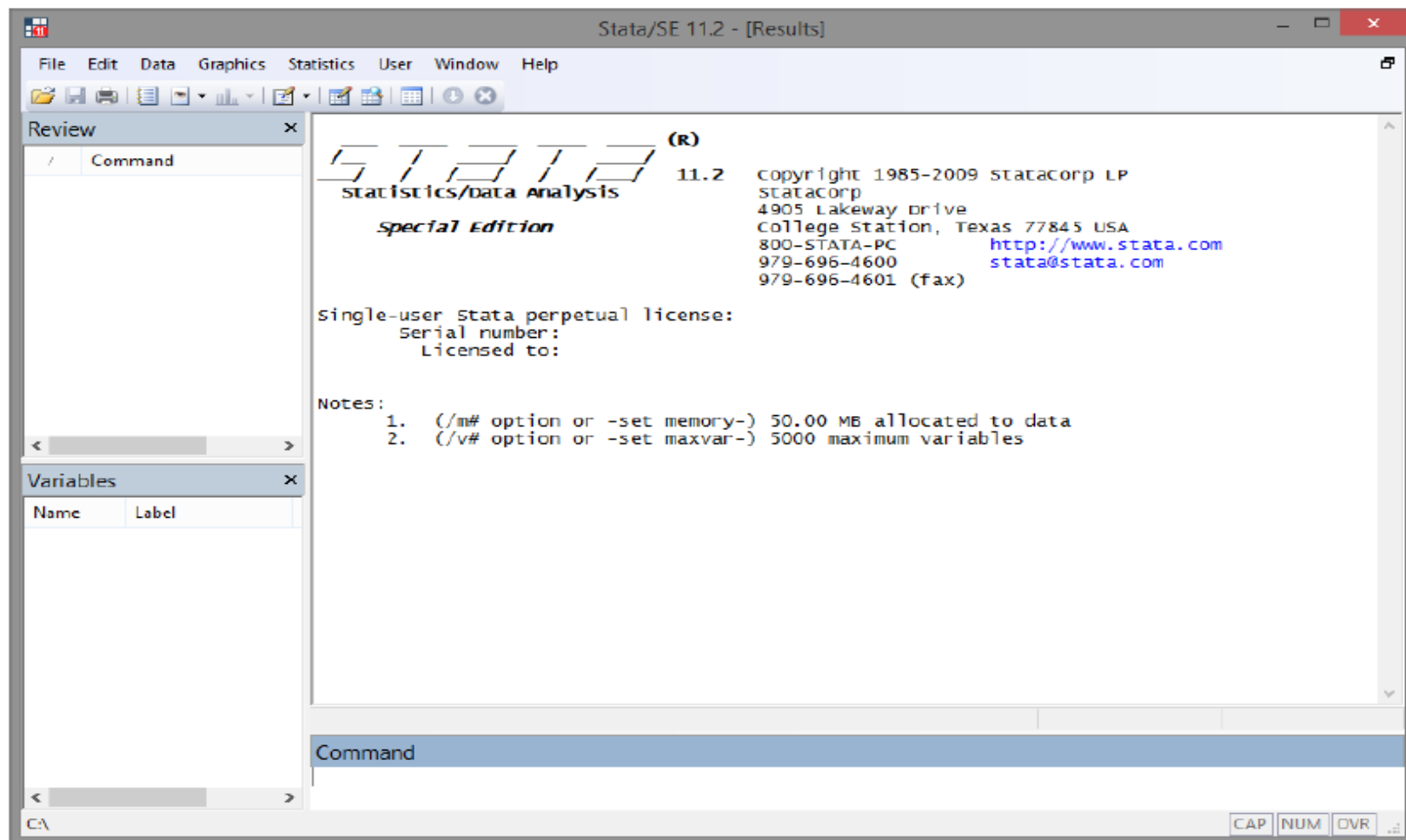
Inmaculada Álvarez Ayuso

Office 314 (Módulo I) www.uam.es/inmaculada.alvarez

E-mail: inmaculada.alvarez@uam.es

Introduction to the use of Stata

Stata 11 Interface



Introduction to the use of Stata

¿How import data?

The file in format .CSV can be imported in Stata using the command **insheet**.

insheet using “name file.csv”, comma (for data separated by comma)

insheet using “name file.csv”, tab (for data separated by tabulated)

insheet using “name file.csv”, delimiter(;) (to especify the option of separation)

Introduction to the use of Stata

Ficheros Excel, .xls y .xlsx

- In versión Stata 12 we can import files in .xls and .xlsx format
- In previous version we can import files in format .CSV with the command `insheet`
- An alternative is the use of the converter programm Stat/Transfer



Introduction to the use of Stata

To save in format .dta

`save` “file name”, `replace`

To download and use in format .dta we can use the command `use`

`use` “file name”

It is convenient to clear before start using data

`clear`

Introduction to the use of Stata

Data base description

Using the command

```
. describe
```

Contains data from **E:\fproduccion.dta**

obs: **551**

vars: **6**

size: **11,020** (99.9% of memory free)

7 Jun 2014 12:41

variable name	storage type	display format	value label	variable label
region	byte	%8.0g		REGION
year	int	%8.0g		YEAR
lngdp	float	%9.0g		LNGDP
time	byte	%8.0g		TIME
lnlabour	float	%9.0g		LNLABOUR
lncapital	float	%9.0g		LNCAPITAL

Sorted by:

Introduction to the use of Stata

Descriptives statistics

We can obtain descriptive statistics using the command `summarize`, or with `sum`

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
region	551	10	5.482203	1	19
year	551	1984	8.374203	1970	1998
ln_gdp	551	8.604461	1.380974	5.008489	10.93575
time	551	15	8.374203	1	29
lnlabour	551	12.83432	1.35033	9.489902	14.70181
lncapital	551	16.29185	1.451554	12.56111	18.63768

Introduction to the use of Stata

Descriptive statistics

To obtain descriptive statistics in detail we can use the command `summarize` «name var», `detail`

```
. summarize lngdp, detail
```

LNGDP				
	Percentiles	Smallest		
1%	5.18547	5.008489		
5%	5.447562	5.032355		
10%	5.896124	5.111677	obs	551
25%	8.036946	5.150204	Sum of wgt.	551
50%	8.723713		Mean	8.604461
		Largest	Std. Dev.	1.380974
75%	9.576183	10.83157		
90%	10.26649	10.84789	Variance	1.907089
95%	10.49854	10.88805	Skewness	-.8685156
99%	10.79785	10.93575	Kurtosis	3.445063

Introduction to the use of Stata

Expression if

Allow to use command for subsamples

For example, `summarize` var 1 for observations after the year yr 1980

```
summarize var 1 if yr >1980
```

Operations that we can use:

Equal (==) and non equal (>, <, >=, <=)

Binary and (&) or (|)

```
summarize var 1 if yr >= 1980 & yr < 1990
```



Introduction to the use of Stata

To generate variables

We can construct variables using the command **generate**, or **gen**

generate «new variable» = expression

For example, an squared variable:

gen var2=var^2

Logaritmic transformation:

gen lvar=log(var)

We can use all operators to generate new variables

Introduction to the use of Stata

To construct variables (cont.)

To generate variables based in special functions `egen`

`egen` «new variable» = expression

For example, to generate a variable which is the mean of other:

`egen varmean=mean(var)`

Or to standarized:

`egen varstd=std(var)`

Introduction to the use of Stata

To generate changes in a variable

To generate changes in a variable we can choose the command **replace**.

If we use the command **generate** and the expression **if** we can create a dummy

For example, a dummy for last years:

```
gen lastyr=0
```

```
replace lastyr = 1 if yr > 1980
```

We can rename with the command **rename**:

```
rename «old variable» «new variable»
```

Introduction to the use of Stata

To repeat a command by subsamples

Using the prefix **by** we can repeat command of stata by subsamples.

by «variable» : command

Before using **by** we need to order the data with **sort**, or using **bysort**, that sort the data

For example, we can use **by** to generate the mean of a variable for years:

```
sort yr
```

```
by yr: egen imeanyr = mean(i)
```

Introduction to the use of Stata

Stored results

Stata commands are classified as being

- r-class: general commands that store results in r()
- e-class: estimation commands that store results in e()
- s-class: parsing commands that store results in s()
- n-class: commands that do not store in r(), e() or s()

Using the syntax `return`, `ereturn` and `sreturn list`, we can list the results stored in r-class, e-class and s-class commands.

Introduction to the use of Stata

Looping while

A very simple command but with wide application. Can be used to effectively reduce anfractuous repetition in do files and programs.

```
while expression {  
    stata commands  
    ...  
}
```

- **while** evaluates the *expression* and if true executes the commands in brackets. It then repeats until *expression* is false
- **while** may be nested within other while commands

Introduction to the use of Stata

Looping while

Example 1

```
local i =1
while `i' < 20    {
display `i'
egen meanyear=mean(year[`i'])
display as text "average year = " as result meanyear
drop meanyear
local i = `i' + 1
}
```


Introduction to the use of Stata

Looping while

Example 2

```
local i = 1970
while `i' < 1971 {
  display `i'
  egen meanIncipital = mean(Incipital) if year == `i'
  display as text "average annual Incipital = " as result meanIncipital
  drop meanIncipital
  local i = `i' + 1
}
```

Introduction to the use of Stata

Looping while

Example 3

```
local i =1970
while `i' < 1979 {
display `i'
twoway (scatter Incapital region ) if year==`i'
local i = `i' + 1
}
```

* This is equivalent to.....

```
twoway (scatter Incapital region ) if year==1970
twoway (scatter Incapital region ) if year==1971
twoway (scatter Incapital region ) if year==1972
twoway (scatter Incapital region ) if year==1973
twoway (scatter Incapital region ) if year==1974
twoway (scatter Incapital region ) if year==1975
twoway (scatter Incapital region ) if year==1976
twoway (scatter Incapital region ) if year==1977
twoway (scatter Incapital region ) if year==1978
```

Introduction to the use of Stata

Looping Foreach

```
foreach lname { in | of listtype } list {  
    Stata commands referring to 'lname'  
}
```

Allowed are

```
foreach lname in any_list {
```

```
foreach lname of local lname {
```

```
foreach lname of global lname {
```

```
foreach lname of varlist varlist {
```

```
foreach lname of newlist newvarlist {
```

```
foreach lname of numlist numlist {
```



Introduction to the use of Stata

Looping Foreach

Example

```
foreach var of varlist region-Incapital {  
  quietly summarize `var'  
  summarize `var' if `var' > r(mean)  
}
```

Introduction to the use of Stata

Looping Forvalues

```
forvalues lname = range {  
    Stata commands referring to 'lname'  
}
```

where *range* is

$\#_1 (\#_d) \#_2$ meaning $\#_1$ to $\#_2$ in steps of $\#_d$

$\#_1 / \#_2$ meaning $\#_1$ to $\#_2$ in steps of 1

$\#_1 \#_t$ to $\#_2$ meaning $\#_1$ to $\#_2$ in steps of $\#_t - \#_1$

$\#_1 \#_t : \#_2$ meaning $\#_1$ to $\#_2$ in steps of $\#_t - \#_1$

The loop is executed as long as calculated values of '*lname*' are $< \#_2$, assuming that $\#_d > 0$.

Introduction to the use of Stata

Looping Forvalues

Example 1:

```
gen x1=region
gen x2=lnlabour
gen x3=lncapital

forvalues num=2/3 {
  summarize x`num'
  egen meanx=mean(x`num') if (x`num'>=13 & x`num'<=15)
  display as text "average value = " as result meanx
  drop meanx
  display "number of observations between 13 and 15: "
  count if (x`num'>=13 & x`num'<=15)
}
```

Introduction to the use of Stata

Looping Forvalues

Example 2:

lists the numbers 1–5, stepping by 1, whereas

```
. forvalues i = 10(-2)1 {  
  2.           display 'i'  
  3. }
```

10

8

6

4

2

Introduction to the use of Stata

Basic graphs

1. Graphs for one continuous variable

- Histogram

- Box-plot

- Pie chart (to represent portions)

- Bar chart (for repeated values)

2. Graphs for two continuous variables

- Scatter plot

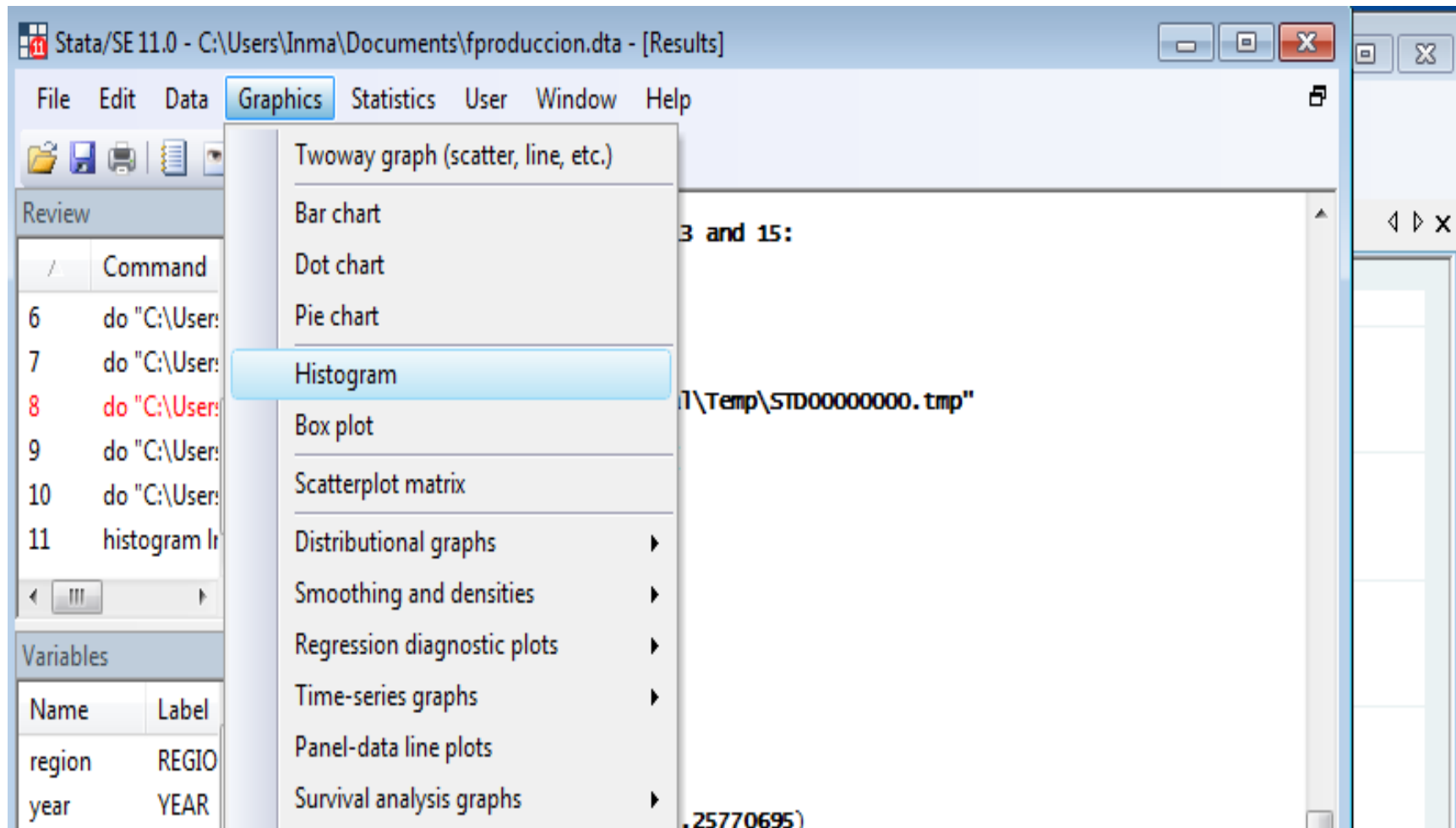
- Fit plot

3. Edit graphs

4. Export graphs

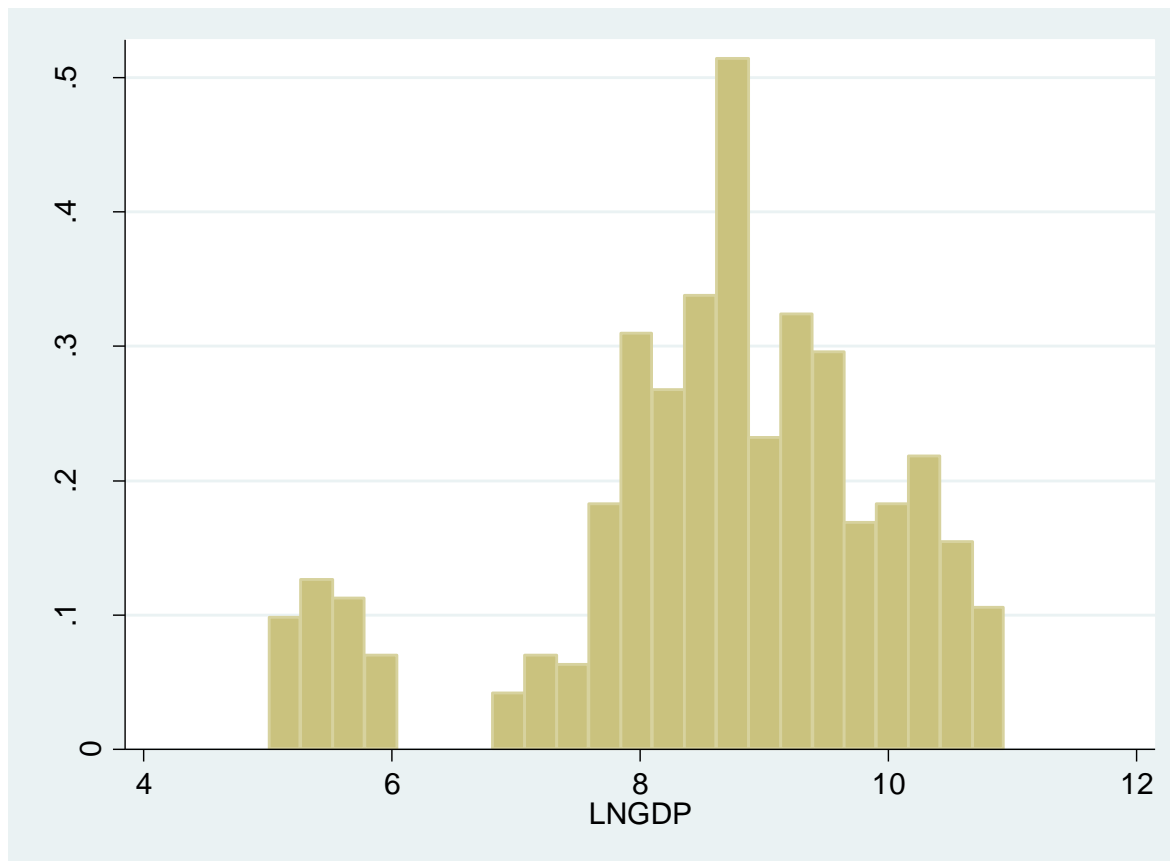
Introduction to the use of Stata

Histogram



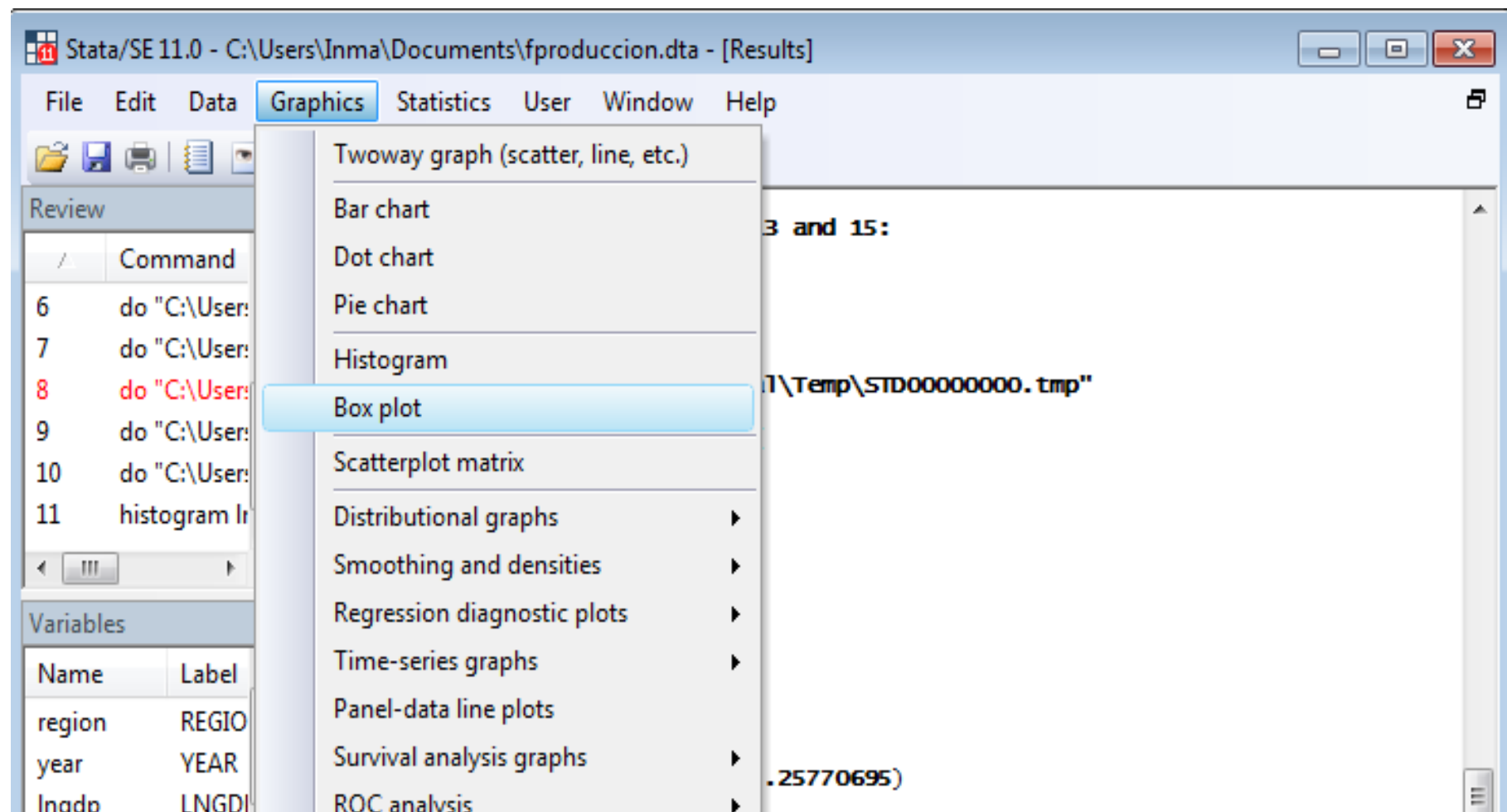
Introduction to the use of Stata

Histogram



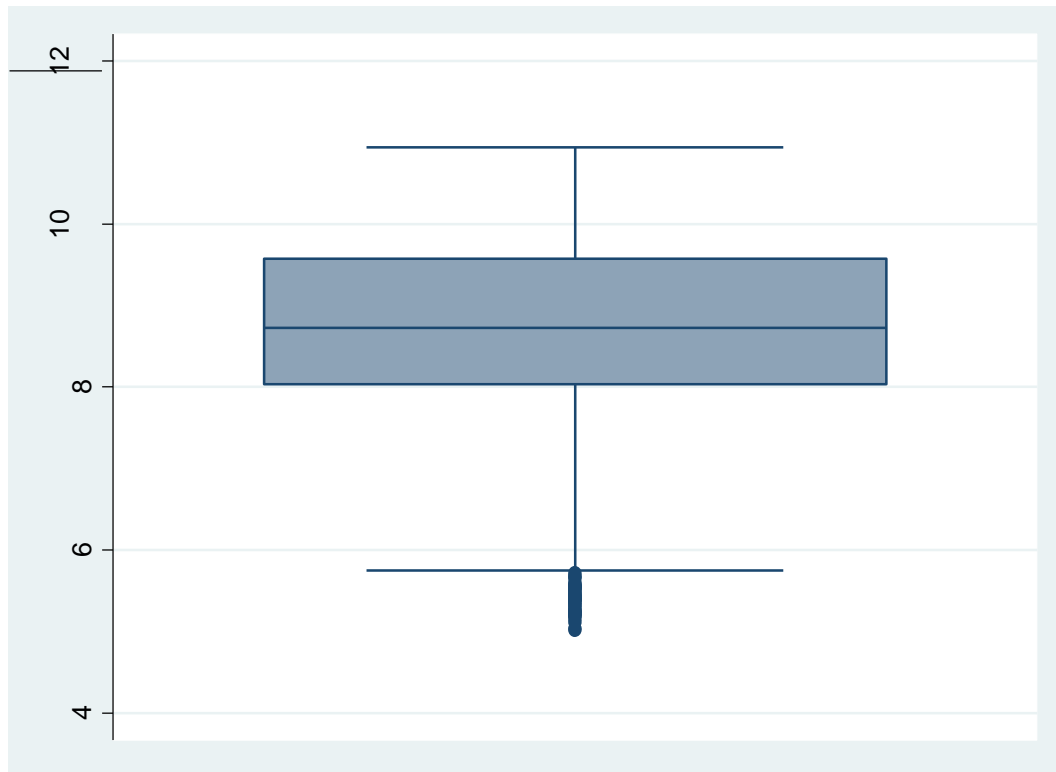
Introduction to the use of Stata

Box-plot



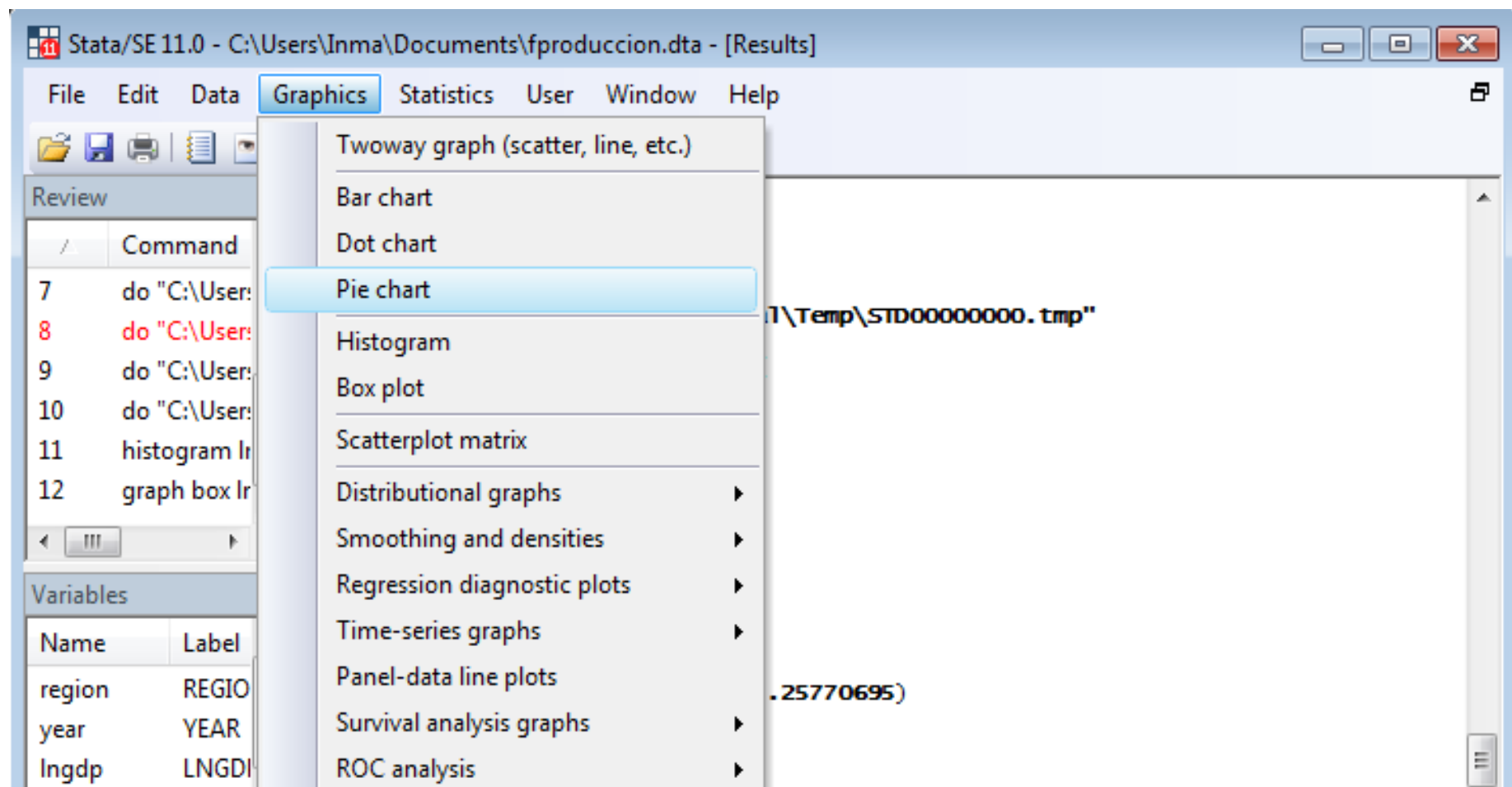
Introduction to the use of Stata

Box-plot



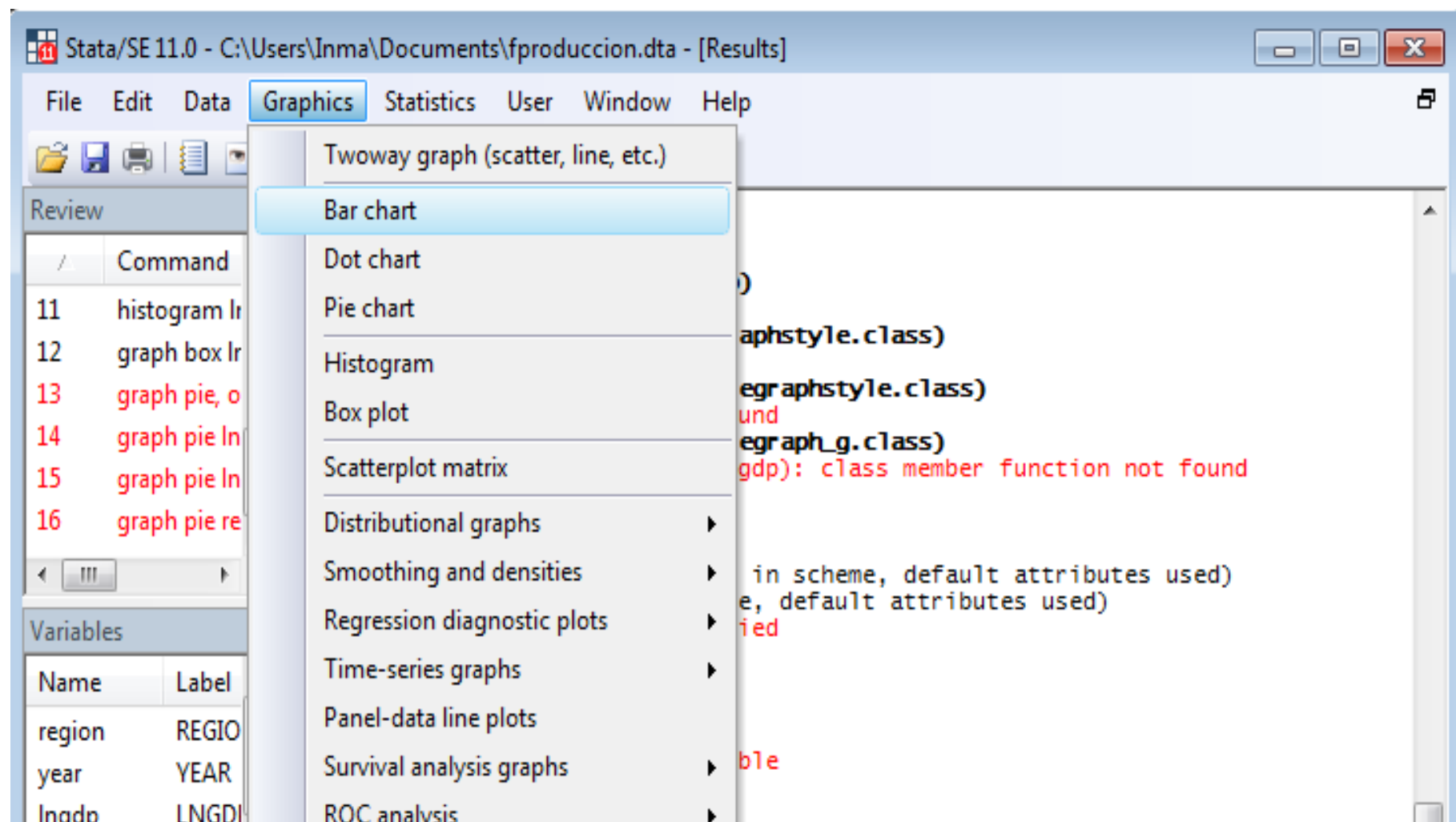
Introduction to the use of Stata

Pie chart



Introduction to the use of Stata

Bar chart



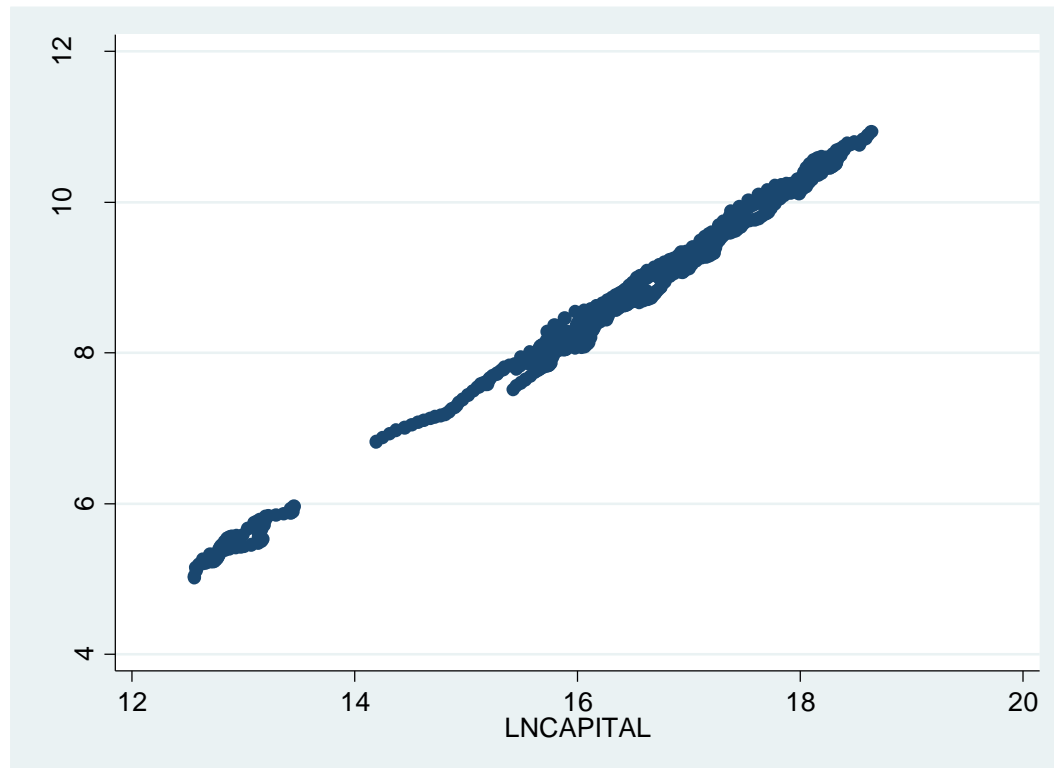
The screenshot shows the Stata/SE 11.0 interface. The title bar reads "Stata/SE 11.0 - C:\Users\Inma\Documents\fprouduccion.dta - [Results]". The menu bar includes File, Edit, Data, Graphics, Statistics, User, Window, and Help. The Graphics menu is open, displaying options: Twoway graph (scatter, line, etc.), Bar chart (highlighted), Dot chart, Pie chart, Histogram, Box plot, Scatterplot matrix, Distributional graphs, Smoothing and densities, Regression diagnostic plots, Time-series graphs, Panel-data line plots, Survival analysis graphs, and ROC analysis. The Command window on the left shows a list of commands: histogram, graph box, graph pie, graph pie ln, graph pie ln, and graph pie re. The Variables window on the left lists: region (REGIO), year (YEAR), and lngdi (LNGDI). The main window displays a command: `graph pie, o` followed by a red error message: `gdp): class member function not found`.

Name	Label
region	REGIO
year	YEAR
lngdi	LNGDI

Introduction to the use of Stata

Scatter plot

```
. twoway (scatter lngdp lncapital)
```



Introduction to the use of Stata

Fit plot

. twoway (lfit lngdp lncapital)

