

An Introduction To Stata and Matlab

Liugang Sheng

ECN 240A

UC Davis

Stata and Matlab in our Lab

- Go to the admin webpage
- <http://admin.econ.ucdavis.edu/Computing/>
- Follow the instruction

http://admin.econ.ucdavis.edu/Computing/TS_Windows/TS-Windows.html

Overview of Stata

- Windows
- Toolbar
- Do file
- Log file
- Data editor
- Version

Loading your data into Stata

- If your data is in STATA format, ie “filename.dta”, then enter:
 use “filename.dta”
- If your data is a comma delimited file:
 insheet using “filename.txt”
- If your data has fixed format or dictionary, find help for
 “infix” and “infile”
- If your data is in excel, you can copy and paste into data
 editor windows
- For other formats, can use the additional software Transfer to
 convert to STATA format

Types of command

- Administrative commands that manage Stata environment, i.e., set memory, matsize, change directory, log file.
- Data management commands that tell stata to read and modify your data
- Commands that tell Stata to carry out the statistical analysis
- Commands that tell Stata to program, Mata environment
- Use help if you have troubles with some commands, and read their descriptions carefully
- Use “findit” to download new commands or packages developed by other programmers. For example

findit xtabond

Example: exp.do

- Shows a general frame of a do file
- Some typical administrative commands
- Shows some useful command for summary statistics
 - list
 - describe
 - codebook
 - summary
 - tabulate

Manage your data: example1.do

- Combining datasets: merge

```
cd "C:\Users\lgsheng\Dropbox\ecn240A"  
use ecn200a.dta,clear  
sort id  
save ecn200a.dta,replace  
use experiment.dta,clear  
sort id  
merge id using ecn200a.dta
```

- Get aggregate level summary statistics

```
collapse (mean) score ecn200a, by(head)
```

Lagged and forwarded variables

- “[_n-1]” tells STATA this is the previous observation
- “[_n-2]” is 2 observations before
- “[_n+1]” is 1 observations ahead
- Examples: (assuming data is sorted by year)
 - `gen GDP_lagged= GDP[_n-1]`
 - `gen GDP_2= GDP[_n-2]`
 - `gen GDP_f1=GDP[_n+1]`

Loops

- “forvalues”

* Generate 100 uniform random variables named x1, x2, ..., x100:

```
forvalues i = 1(1)100 {  
  generate x`i' = uniform()  
}
```

- “foreach”

```
global money "Franc Dollar Lira Pound"  
foreach y of global money {  
  display "`y'"  
}
```

Exercise: caschool.dta

- Compute summary statistics
- Plot the scatter plot of
 - student teacher ratio and district average income
 - Average reading score and district average income
 - Average math score and district average income
- Question: does the high student teacher ratio (teachers/students) helps to improve students' performance?
 - Reg testscr str

Matlab

- A program for numerical computation, based on matrix.
- Overview of windows and toolbar
- Some useful tips before we start
 - If a statement is terminated with a semicolon (;), no results will be displayed. Otherwise results will appear before the next prompt.
 - Variable names are case sensitive
 - Variable names must start with a letter
 - Start comments with the symbol (%)
 - Be careful when you do matrix operation, keep the size consistent
 - Pre-specifying the size of a matrix and avoiding loops could increase speed dramatically
 - Use help and debug when you have code errors

Matrix operation

- Generate a matrix

```
a=[1,2,3;4,5,6;7,8,9]
```

```
b=a(:,2:3) % extract column 2 and 3 of the matrix a
```

```
X=rand(2,5)
```

- Matrix operation

Power	\wedge	or	$\cdot \wedge$	$a \wedge b$	or	$a \cdot \wedge b$
-------	----------	----	----------------	--------------	----	--------------------

Multiplication	$*$	or	$\cdot *$	$a * b$	or	$a \cdot * b$
----------------	-----	----	-----------	---------	----	---------------

Division	$/$	or	$\cdot /$	a / b	or	$a \cdot / b$
----------	-----	----	-----------	---------	----	---------------

or	\backslash	or	$\cdot \backslash$	$b \backslash a$	or	$b \cdot \backslash a$
----	--------------	----	--------------------	------------------	----	------------------------

NOTE: \backslash : left division. The operation $a \backslash b$ is effectively the same as $\text{INV}(A) * B$

- The dot(.) operator is an element by element operation, a and b need to be same size or b be scalar

MATLAB Relational Operators

- MATLAB supports six relational operators.

Less Than	<
Less Than or Equal	<=
Greater Than	>
Greater Than or Equal	>=
Equal To	==
Not Equal To	~=

- MATLAB supports three logical operators.

not	~
and	&
or	



MATLAB Repetition Structures

- A for loop in MATLAB

```
for x = 1: 1 : 10
```

```
    % execute these commands
```

```
end
```

- A while loop in MATLAB

```
while x <= 10
```

```
    % execute these commands
```

```
end
```

MATLAB Selection Structures

- An if - elseif - else structure in MATLAB.

Note that elseif is one word.

```
if      expression1          % is true
    % execute these commands
elseif expression2          % is true
    % execute these commands
else
    % the default
    % execute these commands
end
```

Fsolve-solve nonlinear equations

- Solve a problem specified by

$$F(x) = 0$$

- Example:

$$x^2 + y^2 = 25 \text{ and } x/y = 3/4$$

```
% put it into a M file
function F=myfun2(x)
F(1)=x(1)^2+x(2)^2-25
F(2)=x(1)/x(2)-3/4

g=fsolve(@myfun2,[10 10])
```


Define your own function

- Write your function in a M file
- Example, generate 1000 random draws from Normal distribution with mean 10 and standard variance 5;

```
mu=10; sigma=5; n=1000;
```

```
x=randnorm(mu,sigma,n);
```

```
hist(x,100)
```

- Where the function “randnorm” is a M file;

```
function x=randnorm(mu,sigma,n)
```

```
% this function is to generate n draws of random sample from
```

```
% normal(mu,sigma)
```

```
x=randn(n,1)*sigma+mu
```

- Use debug to find code errors in functions

Plot

- Example

```
x = -pi:.1:pi;
```

```
y = sin(x);
```

```
plot(x,y)
```