

USING STATA FOR DATA WORK

INTRODUCTORY SLIDES

James Ng
james.ng@nd.edu
Center for Digital Scholarship
Hesburgh Libraries, University of Notre Dame

library.nd.edu/cds/

workshop URL

CENTER for
DIGITAL
SCHOLARSHIP

- workshop URL:

<http://www3.nd.edu/~jng2/workshop/>

Stata environment

CENTER for
DIGITAL
SCHOLARSHIP

The screenshot shows the Stata/IC 12.0 software interface. The main window displays the Stata logo and version information (12.0), copyright details (1985-2011 StataCorp LP), and contact information for StataCorp. It also shows a 25-user network perpetual license for ND User at the University of Notre Dame. The interface includes a menu bar (File, Edit, Data, Graphics, Statistics, User, Window, Help), a toolbar, and several panels: a Command window at the bottom, a History window on the left, a Results window in the center, and a Variables window on the right. A Properties window is also visible at the bottom right. Red callout boxes with white text and black boxes with white text are used to label specific parts of the interface: 'History' (C), 'Variables' (B), 'Results' (D), and 'Command line interface' (A).

Stata/IC 12.0 - [Results]

File Edit Data Graphics Statistics User Window Help

Review

Command _rc

There are no items to show.

History

C

(R)

Statistics/Data Analysis 12.0 Copyright 1985-2011 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

25-user Stata network perpetual license:
Serial number: 30120547268
Licensed to: ND User
University of Notre Dame

Notes:
running C:\stata12\profile.do ...

Results

D

Command line interface

Command

A

Variables

Variable Label

There are no items to show.

Variables

Variables

Name
Label
Type
Format
Value Label
Notes

Data

Filename
Label
Notes
Variables 0
Observations 0
Size 0
Memory 64M

CAP NUM OVR

C:\data

ways to use Stata

CENTER for
DIGITAL
SCHOLARSHIP

- point & click
- enter commands in command line interface
- enter commands or code in a “do-file” ← do this for extended projects

good habits for every user

- set Stata's "working directory"
 - if dataset is stored in /Volumes/~jng2/workshop/intro, or if you want any files you produce saved there, set that folder as the working directory:

```
cd /Volumes/~jng2/workshop/intro
```

- how do I find out my working directory?

```
pwd
```


loading data into Stata (1)

CENTER for
DIGITAL
SCHOLARSHIP

Command	File Type	File Extension
use	Stata format	.dta (always)
infix	Fixed-format ASCII	.dat, .raw, .fix, or simply nothing
infile (version 1)	Text-delimited ASCII	
infile (version 2)	Fixed-format ASCII, with a “dictionary”	
import delimited	Text-delimited ASCII	
import excel	Excel	.xls, .xlsx

combining datasets

- merging
 - adding variables to existing observations

id	sex
001	M
002	F

data1.dta



id	age
001	21
002	23

data2.dta



id	sex	age
001	M	21
002	F	23

use data1, clear
merge 1:1 id using data2

- appending
 - adding observations to existing variables

id	sex
001	M
002	F

data1.dta



id	sex
003	F
004	M

data3.dta



id	sex
001	M
002	F
003	F
004	M

use data1, clear
append using data3

more on merging (1)

- for each dataset, must know whether identifying variable/s is/are unique
- in the previous example, the identifying variable is id and clearly unique in each dataset (each value of id only occurs once)
- therefore, we performed a 1:1 (one-to-one) merge
- and here's an example of a m:1 (many-to-one) merge

more on merging (2)

- here's an example of a 1:m (one-to-many) merge
- we want to merge the two files based on state
- state is the identifying variable

state	area
IN	36.4
MI	96.7

data1.dta



state	year	gdp
IN	2014	324289
IN	2013	310669
MI	2014	447221
MI	2013	431112

data2.dta



state	year	area	gdp
IN	2014	36.4	324289
IN	2013	36.4	310669
MI	2014	96.7	447221
MI	2013	96.7	431112

use data1, clear
merge 1:m state using data2

more on merging (3)

- to find out whether the identifying variable/s is/are unique, use the `duplicates report` command
- if the identifying variable is unique, there will be no surplus observations reported

using a “do-file”

- send commands to Stata through a batch file (.do)
 - “do-file”
- Stata reads each line as an executable statement
 - To add comments:
 - start a line with an asterisk * or two slashes //

*this is a comment and will be ignored by Stata

- enclose successive lines with /* and */

/*these two lines are comments
and will be ignored by Stata*/

where to go for help

CENTER for
DIGITAL
SCHOLARSHIP

- Stata has an extensive internal help system
 - need help with how to load data?
`help loading data`
 - need help with `regress` command?
`help regress`
- online resources
 - UCLA: <http://www.ats.ucla.edu/stat/stata/> ← HIGHLY recommend
 - Notre Dame Stata guide:
<http://libguides.library.nd.edu/friendly.php?action=82&s=stata>
 - Google search