Stata for Theses

Debby Kermer
Data Services, University Libraries
dkermer@gmu.edu
Resources

• Stata (check out Statalist)
  – http://www.stata.com/support/

• UCLA
  – http://www.ats.ucla.edu/stat/Stata/

• Social Science Computing Cooperative

• Data Services @ GMU
  – http://dataservices.gmu.edu/software/stata/
  – Debby Kermer  dkermer@gmu.edu
My Workspace
Window Arrangement
Logging (doedit)

capture log close

log using logfile.log, replace

cd S:\name

************************

* stuff goes here

************************

log close

Ctrl-L: Select Line

Ctrl-D: Execute (Do)
Comments in Do Files

* This is a *single-line* comment

/* This is a *multi-line* comment */

// This is a *code* comment used alone

sort year // This is a *code* comment

syntax /* *comment */ syntax
Multi-line Commands in Do Files

recode one (1=3) /// ← must have a space before (2=2) /// (3=1), gen(two)

#delimit ;
recode one (1=3) (2=2) (3=1), gen(two) ;
#delimit cr
Save your do file

clear

Redo your work

Look at your log

```plaintext
** STATA II Workshop  
** dak, 2012

*****************************************************************************
*** ANES: Eval of Govt and Society Study ***
*****************************************************************************

clear
log using logfile.log

cd S:\Debby
use rawdata\ANES_EGSS

*****************************************************************************
*** Prepare File ***
ren C1_* *
sort CASEID
drop VERSION
keep if DC_S1 == 1
varcase(*)

*****************************************************************************
*** Sample Descriptives ***
sort der10ac1
by der10ac1: tab derchoice

*****************************************************************************
*** Population Analysis ***
svset [pweight=weight]
tab derchoice
svy: tab derchoice
svy: proportion derchoice

*****************************************************************************
log close
```
Syntax Suggestions

use working/anes, clear
Syntax

\textit{command} \textit{varlist} \textit{qualifiers, options}

\textit{command} 1, 2, or 3 words specifying the task

\textit{varlist} 0, 1, 2 or more variables

\textit{= exp} A math or logical statement to set a value

\textit{if exp} A math or logical statement to limit cases

\textit{, options} A word or selector to alter the command
Wildcards

codebook m?, compact

codebook m??, c

codebook m*, c

regress dv iv_* cov_*

? = 1 character
* = 1+ characters
egen \( x = function(varlist), options \)

rowmean(varlist)
max(var), by(varlist)
rank(var)
anymatch(varlist), values(numlist)
seq()

egen extreme = anymatch(G?), values(1)
egen composite = rowmean(acs??)

sometimes numbers or an expression
Missing Values

. (a period)

```
tab regic if pv < 8 , s(v) missing
replace has_pet = . if q75 == 9
```

**WARNING:**
missing values = a very large positive number

```
gen eligible = 0 replace eligible = 1 if (age >= 18) INCOMPLETE
replace eligible = 1 if (age >= 18) & (age < .)
replace eligible = 1 if (age >= 18) & !missing(age)
```
## Operators

<table>
<thead>
<tr>
<th>Operator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>==</td>
<td>equals (test)</td>
</tr>
<tr>
<td>!=</td>
<td>not equal to</td>
</tr>
<tr>
<td></td>
<td>or</td>
</tr>
<tr>
<td>&amp;</td>
<td>and</td>
</tr>
<tr>
<td>&lt;</td>
<td>less than</td>
</tr>
<tr>
<td>&lt;=</td>
<td>less than or equal to</td>
</tr>
</tbody>
</table>

**gen eligible** = \( (age >= 18) \& (age < .) \)

<table>
<thead>
<tr>
<th></th>
<th>Adult</th>
<th>Child</th>
<th>No Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>gen eligible</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**gen eligible** = \( (age >= 18) \textbf{if} (age < .) \)

<table>
<thead>
<tr>
<th></th>
<th>Adult</th>
<th>Child</th>
<th>No Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>gen eligible</td>
<td>1</td>
<td>0</td>
<td>.</td>
</tr>
</tbody>
</table>
Variable Commands

```plaintext
rename oldvar newvar
drop variables
keep variables
drop if records
keep if records
ren C1_ * *
drop VERSION
ren C1_ * *
drop VERSION
keep if G1 > 0
keep if G2 > 0
keep if G1 > 0 & G2 > 0
recode G? (min/0 = .), copyrest prefix(R)
```
Converting Types

destring  string  →  numeric

tostring  numeric  →  string

encode  string  →  numeric  text  →  labels

decode  numeric  →  string  labels  →  text

help encode

compress
User-Written Stuff

**Commands** and **Functions**

- `findit varcase`

- `varcase varlist`

- `click on URL`

- `click to install`
.ado Files

**ado** = Automatically Loaded **Do** File

List ado commands:

- **ado**

List ado directories:

- **sysdir**
Files & Directories
Create a Project Folder

Go to Drive S:
   cd S:\

Create a new folder with your name
   mkdir yourname

Change the Working Directory (in the lower left)
   cd yourname

Make a subdirectory for the datafiles:
   mkdir rawdata
   mkdir working
Use and Save

use http://dataservices.gmu.edu/files/anes.txt

save rawdata/anes.dta

clear

use rawdata/anes

save working/anes

clear
insheet using http://dataservices.gmu.edu/files/anes_tab.txt

tab
comma
delimiter(“|”)

outsheet using “working/anes.csv”, comma
outsheet first last using “names.csv”, delimiter(“ “)
Aggregating Files

append using file
append using round2, generate(round)

merge 1:1 key using file
merge 1:1 id using source2, generate(source)
Setup Files

**set mem 20m**
- Memory management is not needed in Stata 12+

**local raw_data “data-filename”**
- Sets a variable used later in the syntax
- Look for `raw_data` farther down

**infile** and **dictionary files**
Stata Principle #2

• Don’t lose Data

• Must specify overwriting:
  – Use `replace` to update variables
  – Use `replace` to update everything else

• Must okay losing changes:
  – Use `clear` or `clear` to exit or `use` a new file
Advanced Syntax
Prefix Commands

help prefix

svy : regress varlist
bysort varname : tab varname
stepwise, pr(.2) : regress varlist

Must be sorted

sort f1
by f1 : ... OR bysort f1 : ...
Complex Samples

```
svyset [ pweight = weight ]
  type these brackets!
```

```
tab derchoice
```

```
svy : tab derchoice
```

```
svy : proportion derchoice
```

Descriptives for Surveys

```
tabulate \rightarrow proportion
```

```
summarize \rightarrow mean
```
Advanced Syntax

```
command varlist =exp if in [weight] using file, options
```

- `= exp` A math or logical statement to set a value
- `if exp` A logical statement to limit cases
- `in range` A range of numbers to limit cases: 1/10
- `[weight]` Weights cases based on sampling details
- `using file` Specifies a single file to use
Charts and Tables

Graph Types

Graph Formatting

Using Graphs
http://www.stata.com/support/faqs/graphics/gph/stata-graphs/
or google "stata graph FAQ"

Box plot of two variables by values of categorical variable

Commands to reproduce

webuse bpwide
graph box bp_before bp_after, over(agegrp)
Formatting Graphs

Saving Graphs

`graph save mygraph`

`graph export mygraph.png`

[http://www.ssc.wisc.edu/sscc/pubs/4-23.htm](http://www.ssc.wisc.edu/sscc/pubs/4-23.htm)
### Saving Tables

<table>
<thead>
<tr>
<th></th>
<th>wrkgovt</th>
<th>mean</th>
<th>N</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wrkgovt</td>
<td>mean</td>
<td>N</td>
<td>sd</td>
</tr>
<tr>
<td>2</td>
<td>govern</td>
<td>41.91346</td>
<td>520</td>
<td>13.84174</td>
</tr>
<tr>
<td>3</td>
<td>private</td>
<td>40.67512</td>
<td>2207</td>
<td>14.49085</td>
</tr>
<tr>
<td>4</td>
<td>Total</td>
<td>40.91126</td>
<td>2727</td>
<td>14.37508</td>
</tr>
</tbody>
</table>

---

**Stata/SE 11.1 - S:\GMU\original\gss2010merged_r1.dta - [Results]**

**Book1 - Microsoft Excel**