

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
 - <http://www.ssc.wisc.edu/sscc/pubs/stat.htm>
- Data Services @ GMU
 - <http://dataservices.gmu.edu/software/stata/>
 - **Debby Kermer** dkermer@gmu.edu

My Workspace

The screenshot displays the Stata 12.1 workspace with several windows open:

- Command Window:** Shows the command `. tab Q1` and its output, including a frequency table and a cumulative percentage table.
- Viewer - help recode:** Displays the help text for the `recode` command, including its title, syntax, and basic/full syntax examples.
- Do-file Editor - Labor_stats.do:** Shows a Stata do-file with commands for listing variables, sorting by work hours, and dropping variables and records.
- Review Window:** Lists the commands entered in the command window, with `. use C:\Rep...` highlighted in red.
- Properties Window:** Shows the properties of the selected variable `CASEID`, including its label, type, and format.

Command Window Output:

```
label: Q1
range: [1,9]          units: 1
unique values: 4      missing.: 0/1058

tabulation: Freq.   Numeric   Label
           402     1   Better
           125     2   Worse
           510     3   About the same
           21      9   DK/NA
```

Expect 2011 vs 2010	Freq.	Percent	Cum.
Better	402	38.00	38.00
Worse	125	11.81	49.81
About the same	510	48.20	98.02
DK/NA	21	1.98	100.00
Total	1,058	100.00	

Viewer - help recode:

Title

[D] recode — Recode categorical variables

Syntax

Basic syntax

```
recode varlist (rule) [(rule) ...] [, generate(newvar)]
```

Full syntax

```
recode varlist (erule) [(erule) ...] [if] [in] [, options]
```

Do-file Editor - Labor_stats.do:

```
19 des, short
20 des
21
22 *** List First 100 Employers ***
23 list wrkgovt in 1/100
24
25 *** Bring up Browse Window ***
26 br
27 |
28 *** Sort by Work Hours ***
29 sort hrs1
30
31 *** Drop Variables ***
32 drop hrs2
33
34 *** Drop Records ***
35 keep if hrs1 > 30
36
```

Review Window:

- 1. help infile
- 1. help merge
- 1. help recode
- 2. use C:\Rep... 4
- 2. inspect Q1
- 2. tab Q1
- 2. tab Q1, nolab
- 2. codebook Q1
- 2. help reshape
- 2. doedit
- 2. tab Q1

Properties Window:

Name	Label	Type	Format	Value Label	Notes
CASEID	caseid	long	%12.0g		

Window Arrangement

The screenshot displays the Stata/SE 12.1 interface with the following components:

- Review Panel (Left):** Shows a list of commands and their results. The first command is `_rc`, which results in "There are no items to show."
- Main Window (Center):** Displays the output of the `ataCorp` command, including contact information for StataCorp LP and license details for a student lab perpetual license.
- Variables Panel (Right):** Shows a table with columns for Variable and Label, currently empty.
- Properties Panel (Right):** Shows a table with columns for Name, Label, Type, Format, Value Label, and Notes, currently empty.
- Data Panel (Right):** Shows a table with columns for Filename, Label, Notes, Variables, Observations, Size, and Memory, currently empty.

The interface includes a menu bar (File, Edit, Data, Graphics, Statistics, User, Window, Help) and a toolbar with various icons. A mouse cursor is visible over a navigation button at the bottom center of the main window.

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)
```

```
#delimiter ;
```

```
recode one (1=3)  
          (2=2)  
          (3=1), gen(two) ;
```

```
#delimiter cr
```

Save your do file

clear

Redo your work

Look at your log

```
32701-0001-Setup.do*  Untitled.do*
1  ** STATA II Workshop
2  ** dak, 2012
3
4  ****
5  *** ANES: Eval of Govt and Society Study ***
6
7  *clear
8  log using logfile.log
9
10 cd S:\Debby
11 use rawdata\ANES_EGSS
12
13 ****
14 *** Prepare File ***
15 ren C1_ * *
16 sort CASEID
17 drop VERSION
18 keep if DC_S1 == 1
19 varcase(*)
20
21 *** Sample Descriptives ***
22 sort der10ac1
23 by der10ac1: tab derchoice
24
25 *** Population Analysis ***
26 svset [pweight=weight]
27
28 tab derchoice
29 svy: tab derchoice
30 svy: proportion derchoice
31
32 ****
33 log close
34
```


Syntax Suggestions

use working/anes, **clear**

Syntax

command ***varlist*** ***qualifiers***, ***options***

command 1, 2, or 3 words specifying the task

varlist 0, 1, 2 or more variables

= exp A math or logical statement to set a value

if exp A math or logical statement to limit cases

, options A word or selector to alter the command

Wildcards

codebook m? , ccompact

codebook m?? , c

codebook m* , c

? = 1 character

* = 1+ characters

regress dv iv_* cov_*

EGen

egen *x* = *function(varlist)*, *options*

rowmean(varlist)

max(var), **by**(*varlist*)

rank(var)

anymatch(varlist), **values**(*numlist*)

seq()

sometimes numbers
or an expression



egen *extreme* = *anymatch(G?)*, **values**(1)

egen *composite* = *rowmean(acs??)*

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

Operators

`==` equals (test)

`!=` not equal to

`|` or

`&` and

`<` less than

`<=` less than or equal to

True **1**

False **0**

Adult *Child* *No Age*

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

1

0

0

gen eligible = (age >= 18) **if** (age < .)

1

0

.

Variable Commands

rename *oldvar newvar*

drop *variables*

keep *variables*

drop if *records*

keep if *records*

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	ignores labels
tostring	numeric	→	string	
<u>encode</u>	string	→	numeric	text → labels
<u>decode</u>	numeric	→	string	labels → text

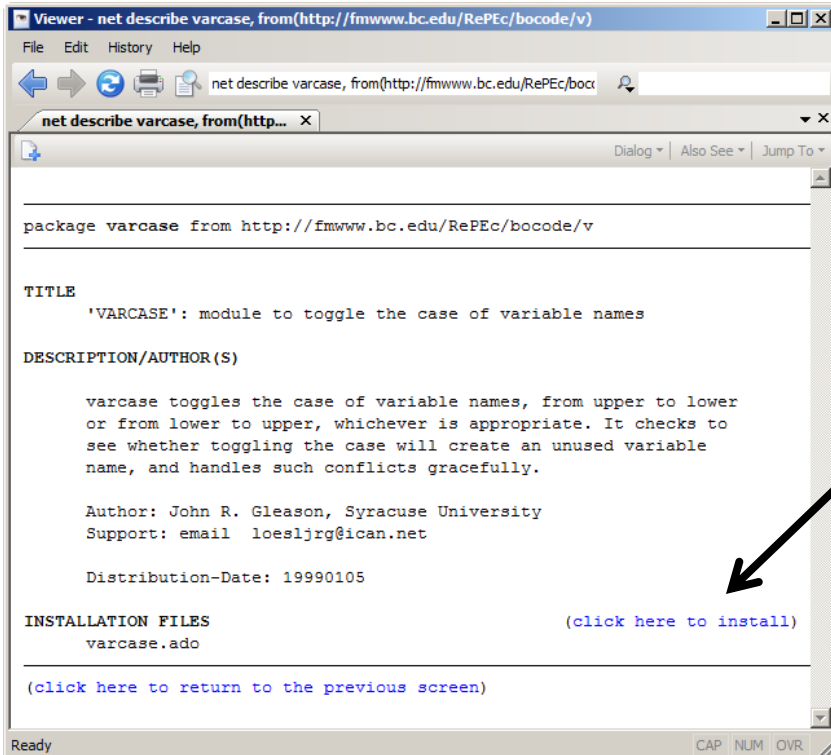
help *encode*

compress

User-Written Stuff

Comands and *Functions*

findit varcase



1 package found (Stata Journal and STB listed first)

[varcase from http://fmwww.bc.edu/RePEc/bocode/v](http://fmwww.bc.edu/RePEc/bocode/v)
'VARCASE': module to toggle the case of variable names, from upper to lower / c whichever is appropriate. It checks to / see whether will create an unused variable / name, and handles

click on URL

click to install

varcase varlist

.ado Files

ado = **A**utomatically Loaded **D**o 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

Delimited Text Files

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+

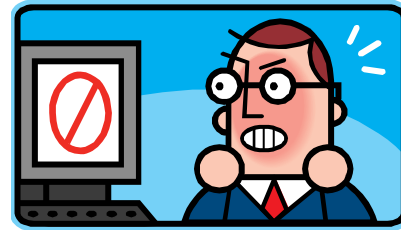
local raw_data “data-filename”

- 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 <i>varlist</i>
bysort <i>varname</i>	:	tab <i>varname</i>
stepwise , pr(.2)	:	regress <i>varlist</i>

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 → **proportion**

summarize → **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

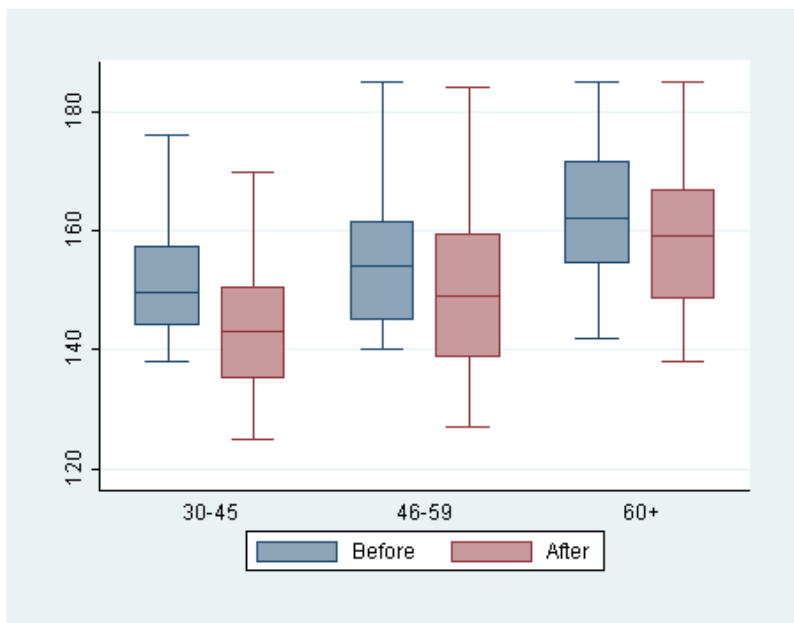
Graph Types

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	Help entries
webuse bpwide graph box bp_before bp_after, over(agegrp)	[G] graph box

[Previous group](#)

[Main page](#)

[Next group](#)

Distribution plots

The grid contains the following plot types:

- Three histograms with overlaid density curves in green, blue, and red.
- Three pairs of box plots in green, orange, and blue.
- Three plots: a symmetry plot (line with points), a gladder plot (two density curves), and a qladder plot (two density curves).
- Three plots: a spike plot (vertical bars), a quantile plot (line with points), and a quantile-normal plot (line with points).
- Three plots: Normal prob (line with points), Chi-squared quant (line with points), and Chi-squared prob (line with points).

Formatting Graphs

- <http://www.ssc.wisc.edu/sscc/pubs/4-24.htm>



Providing Computer Services for the Social Sciences at UW-Madison

• HOME • WINSTAT • WEBMAIL • HELP DESK •

SSCC Knowledge Base

An Introduction to Stata Graphics

Stata includes a rich set of tools for creating publication-quality graphics. Between the wide variety of graphs you can make and the sheer number of details you can control in a graph, Stata graphics can be a daunting subject. However, you'll probably only need to make a few different kinds of graphs, and in most cases Stata's default settings will be fine. What's more, Stata's Graphical User Interface (GUI) organizes the various graphing options in an intuitive way so you can find them when you need them without memorizing the syntax for each one. That doesn't mean you shouldn't put your graph commands in do files once you've created them, but it does mean that for complex graphs you can use the GUI to create the commands you'll store.

Saving Graphs

graph save mygraph

graph export mygraph.png

<http://www.ssc.wisc.edu/sscc/pubs/4-23.htm>



Providing Computer Services for the Social Sciences at UW-Madison

• HOME • WINSTAT • WEBMAIL • HELP DESK •

SSCC Knowledge Base

Using Stata Graphs in Documents

Stata can easily produce "publication quality" graphics, but it takes a few more steps to actually put them in a publication. This article will discuss how to save Stata graphs in various formats. It will also show you step-by-step how to insert a Stata graph into a Word document or PDF file.

Saving Tables

The image shows a Stata window in the background with a Microsoft Excel window overlaid on top. The Excel window displays a table with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	wrkgovt	mean	N	sd								
2												
3	governme	41.91346	520	13.84174								
4	private	40.67512	2207	14.49085								
5												
6	Total	40.91126	2727	14.37508								
7												
8												
9												
10												
11												
12												
13												
14												
15												

The Stata window in the background shows the command window with the following text:

```
Stata/SE 11.1 - S:\GMU\original\gss2010merged_r1.dta - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
A1 wrkgovt
va1]
4852
7462
```