

Stata Quick Guide

prepared by Christine Murray, Van Pelt Library
chrmur@upenn.edu

Data Editor

Allows you to view and edit data in spreadsheet-like interface.

Results window

Displays output of commands. Use **log** command to save.

Variables window

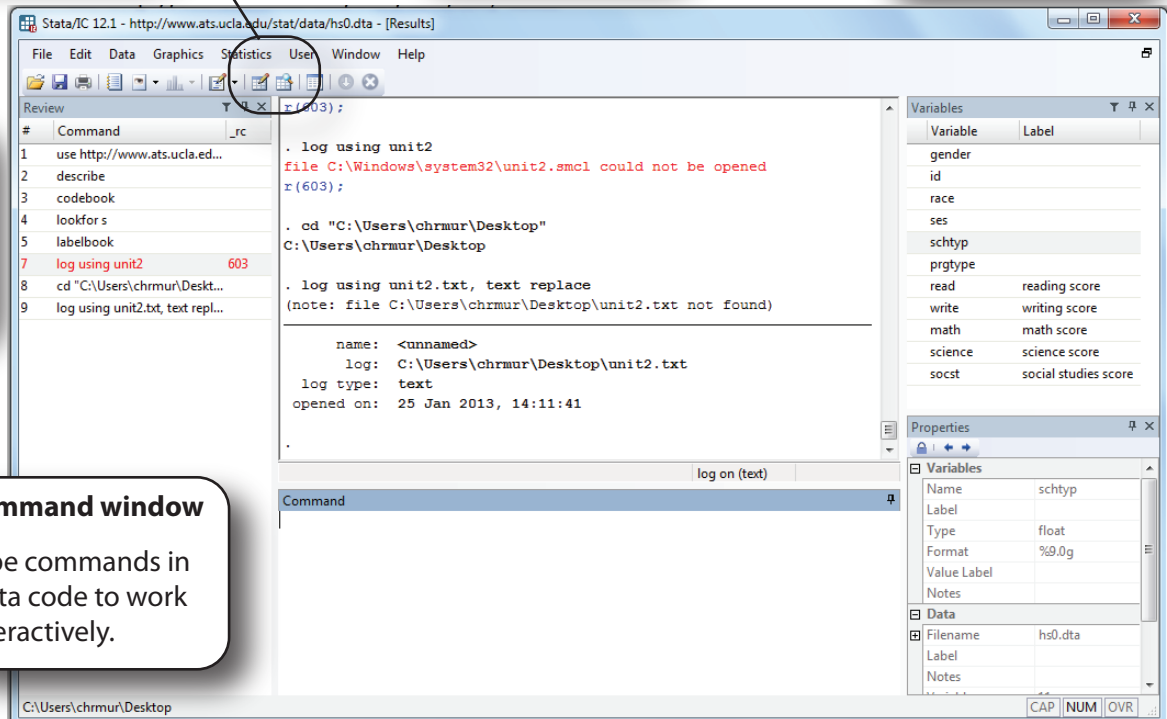
Displays currently loaded variables. Clicking adds to Command window.

Review window

Displays commands already executed. Right-click to save to .do file.

Command window

Type commands in Stata code to work interactively.



Basics

Stata code is generally structured as **command** **variables**, **options**

`bysort variable:` command applies the command to subsets of the dataset according to the variable's values

Underlined portions of commands can be used as abbreviations, as in `su` for `summarize`. Elements in bold vary according to your dataset.

`help command` will access the documentation for that command

A .do file is a series of commands that can be run at once

Loading and Saving Files

`use filename, clear`
open Stata .dta files. **clear** removes currently open data from memory.

`insheet "filename.txt", clear`
load data in spreadsheet format

`save filename, replace`
overwrite and save in .dta format

`log using filename, replace`
saves output in Results window to file

`set memory 50m`
manually sets the memory available to store dataset

Labeling

`label filename "My data"`
a label for the entire dataset

`label variable var1 "My variable"`
label variables

`label define yesno 0 "no" 1 "yes"`
creates label that can be later applied to values of multiple variables

`label values var1 yesno`
applies defined label to variable's values

Exploring Data

`summarize`
descriptive statistics for each variable

`codebook`
prints information on each variable

`lookfor pov`
searches large files for variable containing text, e.g. poverty

`tabulate var1`
one-way tabulation of frequencies

`tabulate var1 var2`
two-way cross-tabulations

`bysort var2: ta var1`
tabulates for each value of `var2`

`histogram var1`
creates histogram graphic

Logical Operators

<code>==</code>	equals	<code>></code>	greater than
<code>!=</code>	not equal	<code>>=</code>	greater than or equal
<code>&</code>	and	<code><</code>	less than
<code> </code>	or	<code><=</code>	less than or equal to
<code>~</code>	not		

Subsetting

`use var1 var2 using filename,`
`clear`
subset by loading only selected variables

`drop var1 or keep var1`
remove or retain selected variables

`keep if var1 > 50`
retain observations according to an expression

Manipulating Variables

`recode var1 -99=. 1/2=1 3/5=2 6=3,`
`gen(newvar)`
creates new variable from existing one, while reassigning values. `""` indicates a missing value in Stata; and `"/"` indicates a range of values

`generate newvar = var1^2`
creates a new variable according to expression

`replace oldvar = var1^2 if var1>10`
alters values in existing variable

`tab var1, gen(dummy)`
creates a dummy variable (0 or 1) for every value of variable

`var*`
selects all variables beginning with text, e.g. `var1, var2, var3, etc.`

Combining Datasets

`sort var1`
sort on values of variable

`merge var1 using otherfile`
adds new columns to currently open data from another file (called the "using data"); both files must be sorted on a variable they have in common

`tab _merge`
check results of merge

`append using otherfile`
adds additional rows of data from another file; must have same variable names