

Vera E. Troeger
16 May 2006

Stata commands:

The structure of stata commands, albeit somewhat arbitrary, follows some logic, namely:

Command (refers to an ado-file)
Variables
Conditions (if...)
, options

You find help to all aspects of the command structure by typing

“help command”

where you have to replace ‘command’ by the name of the ado file you need help on. Getting help in Stata if you don’t know the name of the ado file is somewhat tricky and can be time consuming.

Commands	Options	Description
Before estimation		
set mem	set mem 10000	Increases the memory Stata uses, smaller values are more efficient and quicker, larger values are required as the size of data sets can be large
set matsize	Set mat 800	Defines the size of the Stata matrix,
set more off/on		tells Stata whether to rush through all commands or to make frequent stops, the latter option can be useful if you want to check the results on the screen
cd	cd c:\foldername\foldername	changes the default working directory
use	use <i>filename.dta</i>	opens a dataset
insheet using	insheet using <i>filename.txt</i>	imports a tab-delimited data set and transfers it into State format
log using	log using <i>filename.log</i> , option option: replace – overwrites an existing log file of the same name append – adds to an existing log file of the same name.	opens a logfile which saves all commands and results.
log close		log file is closed
clear		closes data set
exit		exits program.
descriptive stats		
summarize (sum)	sum <i>variable-list</i>	computes summary statistics (Nobs, mean,

		standard deviation, min, max) of all variables in var list. Alternatively: if you don't add a varlist, Stata computes summary statistics for all variables in data set
tabstat	tabstat <i>variable-list</i> , statistics(mean, count, sum, min, max, range, sd, sdmean, skewness, kurtosis, median, p1,..., p99)	tabstat equals sum, but is much more comprehensive
tab1	tab1 <i>varlist</i>	count frequency of listed variables
tabulate	tab <i>var1 var2</i>	tab tabulates var1 (line) und var2 (columns).
list	list <i>variable-list</i>	lists the values of the variables in the list
corr	corr <i>varlist</i>	Pearson's correlation of variables listed
graph	graph y-variables x-variable if, connect(. l) symbol(O S T o d p [varname] [_n]) norm ylabel xlabel wobei: O – large circle S – large square T – large triangle o – small circle d – small diamond p – small plus	allows for various graphical representations of variables in varlist.
Data-management		
sort	sort <i>varlist</i>	sorts observations
move	move <i>var1 var2</i>	changes the order of variables in the data set
gen	gen <i>newvar</i> =expression/ function functions: see Stata help	generates new variables
replace	replace <i>varname</i> = <i>varname</i> / function/ expression	replace values of variables by other values, numbers or a function
rename	rename <i>varname varname</i>	renames a variable.
tsset	tsset <i>panelvar timevar</i>	command required for time series and panel data, defines units and periods
xi	xi i. <i>varname</i>	command required for panel data: generates unit dummies and/or period dummies, tsset required beforehand
Estimation		
reg	reg <i>DV IV1 IV2...</i> if, options options: see Stata help	OLS-Regression

ivreg	ivreg <i>DV (EIV1 EIV2..=iv1 iv2...)</i> IV1 IV2... if, options	IV estimation
logit probit ologit	logit <i>DV IV1 IV2...</i> if, options probit <i>DV IV1 IV2...</i> if, options ologit <i>DV IV1 IV2...</i> if, options options	models for limited dependent variables, binary dependent and categorical dependent variables
arch	arch <i>DV IV1 IV2...</i> if, options options:	auto-regressive and conditional heteroscedasticity models
xtreg xtregar xtpcse xttest0 xttesthausman	xtreg <i>DV IV1 IV2...</i> if, options ... options: fe, re	models for panel data, tsset required fixed effects, random effects,
Post-Estimation commands		
predict	predict <i>newvar</i> , options options: for linear regressions: xb – fitted values der DV = default resid – Residuals rstandard – Residuals rstudent – studentised (jackknifed) Residuals cooksd – cook's distance	computes predictions of the model, including out of sample predictions, various types of residuals, conditional probabilities, and so on
test	test <i>exp1=exp2</i> test <i>var1=var2</i> test <i>varname>x</i>	F-Test for linear models