Introduction to Stata – cheat sheet

Day 1, session 1: importing and managing data

Command	Menu location	Notes
set	Only makes sense from the console or do-file	change Stata settings; new users are likely to use "set more off, permanently"
import delimited	File – Import…	there are other import… commands too, like "import excel" and old commands like "infile". for a .dta Stata data file, try "use"
rename	Data – Variables manager	
keep / drop	Data – Variables manager	keep is the opposite of drop; use whichever has the more succinct *varlist* you can also drop/keep rows (observations) by keep if… or keep in… and likewise for drop
tabstat		summarize is another option, but less flexible
count		particularly useful as "count if…"
correlate / spearman		
generate	Data – Create or change data – Create new variable	if a variable already exists, you might need to "drop" first
label…	Data – Data utilities – Label utilities	
replace	Data – Create or change data – Change contents of variable	to change data in an existing variable
encode	Data – Create or change data – Other variable-transformation commands	
decode		
recode		
tostring / destring		
sort / gsort		gsort is more flexible
list		
save		the replace option is important
collapse		
merge		1:1, m:1, 1:m controls the way in which duplicates are dealt with
reshape		

Day 1, session 2: graphics

Command	Menu location	Notes
scatter	Graphics – Twoway graph	
line	Graphics – Twoway graph	
twoway	Graphics – Twoway graph	alongside scatter and line, area and rspike are very useful
histogram	Graphics – Histogram	
graph bar	Graphics – Bar chart	
graph box	Graphics – Box plot	
graph export / graph save	From the graph window	

Day 1, session 3: tests

Command	Menu location	Notes
ttest	Statistics – Summaries, tests and tables – Classical tests of hypotheses – t test	
tabulate	Statistics – Summaries, tests and tables – Two-way table	with the chi2 option
signrank / ranksum / kwallis	Statistics – Summaries, tests and tables – Nonparametric tests of hypotheses	
regress	Statistics – Linear models and related – Linear regression	
predict	Only makes sense in console or do-file	
test		
margins / marginsplot		
estat ic		

Day 1, session 4: programming and exporting

Command	Menu location	Notes
local / global	When automating work by programming techniques, you need to work from do-files.	Call the local macro: `mylocal' Call the global: $myglobal or, if you have to avoid ambiguities: ${myglobal}
forvalues / foreach		
if / else		
display		
levelsof		
program…		This is a whole collection of commands that we'll just signpost briefly
capture		

dyntext		
dyndoc		
webdoc / tabout / ht / htmlutil / html-reports		User-written commands; have to be installed from SSC or Github.

Day 2, session 1: time series

Command	Menu location	Notes
tsset		
date		
day, week, year, etc		
format		
L, D, F and S (operators)		
tssmooth		
tsline		
ac, pac		
arima, estat ic, wntestq, wntestb		
arch [, garch()]		
varsoc, varbasic, fcast		
dfuller		
itsa		User-written command; has to be installed using **ssc install**

Day 2, session 2: more complex regression

Command	Menu location	Notes
glm / logistic / poisson / nbreg	Statistics – Binary outcomes / Count outcomes / Generalized Linear Models	
marginsplot	Only makes sense in console or do-file	
mixed		
melogit etc		
npregress		

Day 2, session 3: survival analysis

Command	Menu location	Notes
stset	Statistics – Survival analysis – Setup and utilities – Declare data to be…	Has to be done first, like tsset. Can be changed or removed later; does not alter the data.
strate	Statistics – Survival analysis – Summary statistics… - Tabulate failure rates…	

streg	Statistics – Survival analysis – Regression models – Parametric regression models	
stcox	Statistics – Survival analysis – Regression models – Cox proportional hazards models	
sts graph	Statistics – Survival analysis – Graphs	

Day 2, session 4: high-dimensional / multivariate analysis and integrating with R

Command	Menu location	Notes
pca / screeplot	Statistics – Multivariate analysis – Factor and Principal Component Analysis	
mca	Statistics – Multivariate analysis – Correspondence Analysis	
mds	Statistics – Multivariate analysis – Multidimensional Scaling	
cluster kmeans / cluster averagelinkage	Statistics – Multivariate analysis – Cluster analysis	
cluster dendrogram		
rdump		User-written command (provided with this course)
block		User-written command (available at github.com sergiocorreia/stata-misc)
file open / file write / file close	Only makes sense in a do-file	
! / shell / windows / windowsmonitor	Only makes sense in a do-file	windowsmonitor is user-written and on SSC