Data Processing
with Stata Cheat Sheet

For more info see Stata's reference manual (stata.com)

Reading a help file

[**by** *varlist***:**] *command* [*varlist*] [**=***exp*] [**if** *exp*] [**in** *range*] [*weight*] [**using** *filename*] [**,***options*]

ex: by rep78: summarize price if foreign == 0 & price <= 9000

Basic Data Operations

Arithmetic		Logic			
+	add (numbers) concatenate (strings)	**&**	and	**<**	less than
−	subtract	**!** or **~**	not	**<=**	less than or equal
*	multiply	**|**	or	**>**	greater than
/	divide	**==**	equal	**>=**	greater or equal
^	raise to a power	**!=** or **~=**	not equal to		

Set up

pwd
print current (working) directory

cd "C:\Program Files (x86)\Stata13"
change working drive

dir
display filenames in working directory

fs
lists the names of files in compact form*

fs *.dta
List all Stata files in working directory

clear
delete data set in memory

mkdir "C:\MyStataDirectory"
Create a new directory named MyStataDirectory

capture log close
Close the log on any existing do files

log using "$pathlog/myDoFile.do", **replace**
Create a new log file to record your work and results

Importing Data

Viewing & Summarizing Data

use auto2.dta, clear
Load the auto dataset

use make price mpg using auto2.dta, clear
Load a subset of data from a file

sysuse auto2, clear
Load system data (Auto data)

webuse auto2.dta
Load data from the web

Selecting Columns

count

drop make
drop the make variable from auto.dta

Filtering Rows

drop if mpg < 20
drop observations based on a condition

drop in 1/4
drop the first four observations of a dataset

keep in 1/30
Same syntax as drop

sort price mpg
sort the data in ascending order by price then miles per gallon

gsort -price -mpg
Sort in descending order, first by price then miles per gallon

bysort rep78: tabulate foreign
repeat command on subset of data

generate mpgSquared = mpg * mpg
Create or change contents of a variable

save "C:/MyStataDirectory/auto_edit.dta", **replace**
Save data, replace if a file with same name exists

egen unique_ID = group(var1 var2...)
Create a unique id from a combination of variables

assert rep78 != .
Verify the truth of a claim

expand
Duplicate observations

codebook
Describe contents of data file

Creating Data

Transforming Data

Reshaping Data

Tim Essam (tessam@usaid.gov) • Laura Hughes (lhughes@usaid.gov)
Disclaimer: we are not affiliated with Stata. But we like it.

inspired by RStudio's awesome Cheat Sheets (rstudio.com/resources/cheatsheets)   geocenter.github.io/StataTraining   updated January 2016