

Stata Data Management Workshop

Command	Description	Sample Code	Sample Description
help (command name)	Pulls up STATA's help menu	help gen	Brings up the help menu for "gen"
// comments	Comment at the end of a line	clear all // This is a comment	Uses the command clear all and then adds a comment at the end of the line
* comments	Single line comment	* More comments	Can only be used at the beginning of a line
/* comments */	Anything between /* and */ will be commented out	/* This is a comment ... And so is this */	All of the text is interpreted as a comment by STATA
clear all	Clears everything from STATA's memory	clear all	See description
cd	Change the directory	cd "C:\Downloads"	Tells STATA to open and save files in the "C:\Downloads" folder
use	Reads in STATA (.dta) data	use Entrance, clear	Reads in "Entrance.dta" after clearing old data from memory
import delimited	Reads in delimited (.txt, .dat, .csv) data	import delimited entrance.txt, clear	Reads in the file "entrance.txt" after clearing previous data from memory
import excel	Reads in a spreadsheet (.xlsx) data	import delimited flavor2.csv, clear	Reads in the file "flavor2.csv" after clearing previous data from memory
rename	Rename an existing variable	rename sat SAT	Renames sat "SAT"
label variable	Give an entire variable a label	label var gender "Male"	Adds the label "gender" to the variable Male
label define	Create a label for given values. The name you give the label can be anything.	label define race_label 0"white" 1"black" 2"hispanic" 3 "asian"	Creates a label named "race_label" that will label all values of 0 as "white" and all values of 1 as "black" and so forth.
label values	Apply the created label to values	label values race race_label	Applies the label "race_label" to the variable "race"

Stata Data Management Workshop

tostring	Convert a numeric variable to text (string)	tostring id, replace	Converts the variable id into string variable
destring	Convert a text (string) variable to numeric. All values of the variable must be numbers.	destring id, replace	Converts the variable id into numeric variable
encode	Encode string into numeric and vice versa	encode sex, generate(gender)	Turns the string "sex" variable into a numeric variable called "gender"
gen	Create the contents of variable	gen ImAStudent = 1	Creates a variable called ImAStudent and sets it equal to 1
	Create a dichotomous variable	gen YoungFem = ((Gender==1) & (Age == 18))	Generates a dichotomous variable, where YoungFem = 0 if the statement is false and YoungFem = 1 if the statement is true
egen	Extensions to generate	egen rtotal = rowtotal(a b c)	It creates the sum of the variables "a," "b," and "c", treating missing as 0. If all values in "a," "b," and "c" are missing for an observation, newvar is set to missing.
recode	Changing the values of a numeric variable. Does not work with string variables. Replace can be used for string variables.	recode v4 (2=1) (1=0), generate(Gender)	Gender now equals 0 for females and 1 for males
	Create a new categorical variable using the recode command	recode v2 (17/19=1) (20/max=0), gen(teenager)	New variable teenager = 1 if age value ranges from 17 to 19
drop	Removes variables or observations	drop teenager	Removes the teenager variable
		drop if Flavor ≤ 2	Removes observations if the Flavor variable is less than 2

Stata Data Management Workshop

Tips and Tricks

- Stata files can also be opened by dragging them into Stata
- Stata is case sensitive
 - All commands are lowercase
 - Variable names must match exactly
- Use the keyboard shortcut (control d) to execute commands in the do-file
- It is generally unnecessary to save changes to your data set if you used a do-file
 - The do-file should be saved, and can be re-run to replicate what you already did
 - Any saves should be made with a new file name so as not to change your original file

run to replicate