Introductory Stata 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
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.dta", clear	Reads in "Entrance.dta" after clearing "flavor2.csv" from memory
infile using	Reads in ASCII (.txt) data	infile id gender age gpa act sat actsat athlete using "entrance.txt"	Reads in the file "entrance.txt"
insheet using	Reads in a spreadsheet (.csv) data	insheet using "flavor2.csv", clear	Reads in the file "falvor2.csv" after clearing previous data from memory
clear all	Clears everything from STATA's memory	clear all	See description
browse	Opens the Data Browser. Can be used by itself. Can also enter in variable names to view only specific variables.	browse	See description
edit	Opens the Data Editor. Can be used by itself. Can also enter in variable names to view only specific variables.	edit	See description
summarize	# of obs., mean, std. dev., min, max	sum gpa	Summary info. for gpa

Introductory Stata workshop

sum variable, detail	Summarize command with more details	sum gpa, d	Summary of gpa with percentiles, skewness, kurtosis, median, etc.
mean	Estimates the true mean w/ conf. intervals	mean act	Estimate mean of act
tabulate	Creates two-way frequency tables	tab act gender	Two-way frequency table of act and gender
tab1	Creates two-way frequency tables	tab1 act gender	Two-way frequency table of act and gender
frequency	Lists frequency tables	fre act gender	Separate frequency tables of act and gender
proportion	Estimate proportions with conf. intervals	proportion age	Estimated proportion of age by each unique value of the var. age
histogram	Histogram of a single variable	histogram gpa	Histogram of the variable gpa
scatter	Scatter plot of two variables	scatter gpa act, title(gpa/act Regression)	Scatter plot of gpa vs. act with the specified title
pwcorr	Produces a correlation table	pwcorr gpa act, sig	Shows correlation and significance between gpa and act
reg	Computes a linear regression of the included variables	reg gpa act	Regresses act on gpa
generate	Create a variable	gen fpc = 1	new variable fpc with all cells = 1
	Create a dummy variable	gen YoungFem = ((gender==1) & (age==17))	YoungFem = 1 if gender is 1 and is 17 (= 0 otherwise)
	Create an ID, or serial variable	gen id_2 = _n	Id_2 = the observation # of in the dataset for each observation
	Create a variable = # obs.	gen obs = _N	Obs = 49 , which is the total # of obs. In the data

Introductory Stata workshop

egen	Generate with functions There are a number of functions that can be used with the egen command. Do a "help egen" command for a full list of all functions.	egen actsat_rank = rank(actsat) egen actsat_rank_age = rank(actsat), by(age) egen mean_actsat = mean(actsat) egen mean_actsat_age = mean(actsat), by(age)	Generate the variable actsat_rank, shows the rank of each person on their ACT and SAT scores. Can also look at rank by their age. You can also generate a mean score for a variable, and you can generate a mean score, by another variable.
preserve	Temporarily preserve the current data set. Must be done before making changes to data.	preserve	
restore	Restores data to preserved point		
rename	Rename an existing variable	rename sat SAT	Renames sat "SAT"
recode	Changing the values of a numeric variable. Does not work with string variables. Replace can be used for string variables.	recode gender (2=1) (1=0)	Gender now equals 0 for females and 1 for males
	Create a new categorical variable using the recode command	recode age (17/18=0) (else=1), gen(old)	New variable old = 0 if age value ranges from 17 to 18
drop	Removes variables or observations	drop if gpa < 2.00	Removes observations where gpa less than 2
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"
sort	Sort the entire data by listed variables	sort age	Sorts by the value of age (ascending is the default)

Introductory Stata workshop

order	Rearrange the order of variables in the data	order actsat actsat_rank actsat_rank_age	Places the variables listed in that order in the data set
list	Lists values of variables in the output window	list age in 1/10	Lists the first ten values of age (in the sorted order)
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

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