Tips for Stata, Sociology 38X Courses, and Graduate School

Tips for using Stata

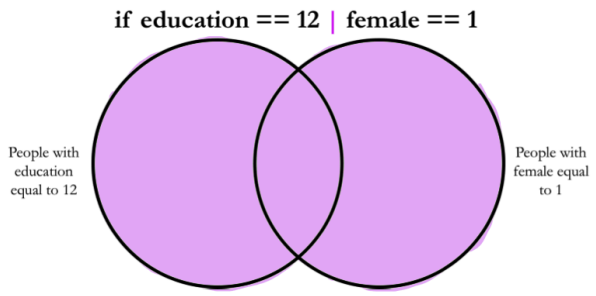
Getting started:

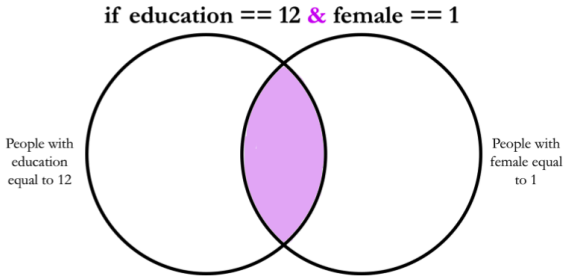
Purpose	**Command**	**Example Code and Description**	
Set the working directory (the folder Stata is accessing on your computer)	cd	cd "Documents/Sociology 381/Homeworks"	→ changes the directory to my folder Documents > Sociology381 > Homeworks
Start a log file	log using *or* log using..., replace	log using "HW1_log.log" *or* log using "HW1_log.log", replace	→ starts a new log and saves it as "HW1_log" → starts a new log and saves it as "HW1_log", replacing any previous log I have by the same name
Close a log file	log close	log close	→ stops the current log file and closes it
Open a data file	use	use "HW1_data.dta"	→ opens "HW1_data"
Save a data file	save *or* save..., replace	save "HW1_data.dta" *or* save "HW1_data.dta", replace	→ saves any changes I made to the data set as "HW1_data" → saves any changes I made to the data set as "HW1_data", replacing any data file I have by the same name
Clear data, etc.	clear	clear *or* clear all *and other options*	→ clears data and value labels → clears data, value labels, saved results, etc.
Describe the data or a variable	describe desc	desc *or* desc income	→ describes all data → describes income variable
Browse the data (i.e., look at the data without changing it)	browse br	br *or* br education	→ opens the data browser → opens the data browser, showing only the education variable
Edit the data (i.e., look at the data *and* change it)	edit ed	ed *or* ed education	→ opens the data editor → opens the data editor, showing only the education variable

Logic statements: In combination with "if," which can be used with most commands (before the comma!)

Statement	Sign	Example Code and Description	
Equal to	==	if education == 12	→ only for those with education equal to 12
Greater than	>	if education > 12	→ only for those with education greater than 12
Less than	<	if education < 12	→ only for those with education less than 12
Greater than or equal to	>=	if education >= 12	→ only for those with education greater than or equal to 12
Less than or equal to	<=	if education <= 12	→ only for those with education less than or equal to 12
Not	~ *or* !	if education != 12 *or* if education ~=12	→ only for those with education not equal to 12
Or	|	if education == 12 | female == 1	→ only for those with education equal to 12 *or* with female equal to 1 (note how this is different from the row below, see illustration below this table)
And	&	if education == 12 & female == 1	→ only for those with education equal to 12 *and* with female equal to 1 (note how this is different from the row above, see illustration below this table)

A short illustration of logical "or" (|) vs. logical "and" (&):





Manipulating and analyzing data:

Purpose	**Command**	**Example Code and Description**	
Get summary statistics (mean, SD) for a variable	summarize sum	sum income	→ summarizes income variable
Tabulate (# and % of every value) one or two variables	tabulate tab	tab gender *or* tab gender education	→ tabulates the gender variable → shows the cross-tabulation of gender and education
Count the number of observations	count	count if income >=20 & income < 40	→ counts the number of observations that have income greater than or equal to 20 and less than 40
Create a new variable	generate gen	gen class = . *or* gen class = 1	→ creates a new variable "class" and sets everyone equal to missing → creates a new variable "class" and sets everyone equal to 1
Replace the values of a variable	replace	replace class = 2 if income >= 20 & income < 40	→ changes the class variable to 2 for everyone with income greater than equal to 20 and less than 40
Recode a variable	recode *or* recode..., generate(...)	recode income (99 = .) *or* recode opinion (1=3) (3=1), gen (reverse_opinion)	→ recodes income variable so those that were equal to 99 are now missing → recodes opinion variable so those that were equal to 1 are now 3 and those that were equal to 3 are now 1, and saves this as a new variable called reverse_opinion (note that those with opinion equal to 2 are still 2 in the new reverse-coded variable)
Destring a variable (useful when numbers are saved as text)	destring..., generate(...) *or* destring..., replace	destring education, gen(education_numeric) *or* destring education, replace	→ changes the education variable so that the numbers are saved as numbers (rather than text), saves it as a new variable "education_numeric" → changes the education variable so that the numbers are saved as numbers (rather than text), replacing the old education variable with the new one
Destring a variable (useful when you want to convert text to numbers and keep the value labels)	encode..., generate(...)	encode income, generate (income_numeric)	→ changes the income variable so that text is now saved as numbers, keeps the text as the value labels, and saves it as a new variable "income_numeric"
Drop (delete) a variable	drop	drop class	→ deletes the class variable
Label a variable	label variable label var	label var class "Social Class"	→ labels the class variable "Social Class"

Define value labels (doesn't apply them yet - see row below)	label define	label define class_lbl 1 "Lower class" 2 "Middle class" 3 "Upper class"	→ creates a new label called class_lbl, where 1 is defined as "Lower class," 2 is "Middle class," and 3 is "Upper class"
Add value labels to a variable (applies value labels from row above)	label values	label values class class_lbl	→ adds the label values defined in class_lbl to the class variable
Sort the data by a variable and run a command separately for each category of that variable	bysort ... :	bysort class: sum income *or* bysort class: tab education	→ summarizes income for each class category → tabulates education for each class category
Create a histogram	histogram hist	hist opinion	→ creates a histogram of the opinion variable
Create a scatterplot	scatter [Y] [X]	scatter income education	→ creates a scatterplot with income as the dependent (Y) variable and education as the independent (X) variable
Calculate an OLS regression	regress [Y] [X] reg [Y] [X]	reg income education female	→ calculates a regression where income is the dependent (Y) variable and education and female are the independent (X) variables

Best practices:

- Always use a do file to save your code!
    - Add comments to your do file to explain your code. Any text following "//" or "*" will be understood by Stata as a comment rather than a command.
- Visualize your data to get a sense of what's going on. Try a histogram or a scatterplot.
- When creating a new variable, tabulate it against the old variable (i.e., tab old_variable new_variable) to make sure it is correct.
- Name new variables, labels, etc. in a way that makes sense to you.
    - Find your balance between length (shorter means faster coding!) and clarity (clarity means no confusion!).
    - Consider naming dummy variables according to whatever value is equal to 1. For example, call a 0/1 variable *female* instead of *sex* if 1 means female .
    - Consider naming recoded variables in a way that helps you identify the original variable. For example, call a variable *education_recode* if it is recoded from *education*.
- Use "if" statements to restrict your sample to only those you want included in the analysis.
    - Use parentheses when making more complicated "if" statements. For instance, "if education == 12 & (class == 2 | class ==3)." This restricts the sample to only those who have an education equal to 12 AND are either in class 2 or class 3.
- Help files: typing "help [command name]" will open the help file for the command, describing the syntax and function of the command (for instance, type "help summarize" for help on obtaining summary statistics for a variable).

- Within the help files, square brackets indicate optional specifications that you are not required to make. For example, "summarize [varlist]" indicates that you can specify a list of variables that you want summary statistics on, but the command will still work if you omit the variable list.
- The order of symbols is important. We usually specify the "varlist" first, followed by "if" clauses, and other optional specifications that are separated by a comma. Pay attention to the placement of the comma, which typically comes after "if" clauses.
- Underlined letters indicate the essential parts of a command; the other letters are optional. For example, "summarize" indicates that we can simply type "su income" to obtain summary statistics for our income variable.

Tips for which statistical technique to use

*(*note: we didn't cover these yet in bootcamp but you will in the methods sequence!)*

Add to this as you advance in the sequence:

Technique	Independent Variable	Dependent Variable	Example
T-test	Binary	Continuous	Wait time in ER (in minutes) by insurance status (0/1)
Chi-square test of independence	Categorical	Categorical	Having been in the ER 72 hours earlier (0/1) by insurance status (0/1)
Correlation	Continuous	Continuous	Length of stay in hospital (days) by age (years)
Ordinary Least Squares (Linear) Regression	Anything	Continuous	Length of stay in hospital (days) on age (years) and/or insurance status (0/1)
Logistic Regression	Anything	Binary	Insurance status (0/1) on sex (0/1), age (years), and/or pain (1-5)

Tips for thriving in the Sociology methods sequence

- Practice using a growth mindset.
  - *There will be things you do not fully understand. That is okay and, in fact, expected. As Tomás says, Stanford is an institute of higher* **learning***, not an institute of higher* **knowing.**
  - Ask questions.
  - Ask for help.

    - Use comments on graded problems sets to improve your next assignment. Ask your TA to explain the comments if you don't understand them.
    - Be patient with yourself.
- Collaborate with your classmates.
    - Schedule times to work as a group (virtually works too!)
    - Leverage each others' strengths--in Stata, statistics, writing, etc.
    - Be patient with your classmates and show appreciation.

Tips for navigating the Sociology department and Stanford as an institution

- Be kind to the administrative staff.
- Double-check your pay stubs.
- Reach out to your advisor to set up meetings; don't just wait for them to check in with you.
- When in doubt:
    - Regarding classes: ask a TA or professor.
    - Regarding Stanford logistics (e.g., about conference travel, getting paid, getting reimbursed): ask
        - Natasha (Graduate Student Services),
        - Randy (Administrative Associate),
        - Mariette (Faculty Affairs and keeper of cords and adapters),
        - Morgan (Undergraduate Student Services and scheduler of discussion sections),
        - Chrissy (Finance Manager and NSF grant expert),
        - or Sihla (Manager).
    - Ask a representative (e.g., Co-Chair Cat) on the Association of Sociology Graduate Students (ASGS).
    - Ask more experienced graduate students (Check out the Slack!)

Tips for thriving in graduate school

- Be kind to yourself.
    - *Remember that you are not your work.*
    - Practice a growth mindset and be reasonable in your expectations for yourself.
    - Continue doing things that energize you and be sure to take time off.
    - Create space away from graduate school. Maintain friendships outside of graduate school. Leave campus every once in a while. Get involved in communities (e.g., student groups) outside of the department.
    - Talk to a therapist.
- Manage your time wisely.
    - Set formative and final deadlines for projects. Use conference submission deadlines to motivate yourself.
    - Schedule time for work and avoid using social media and checking email during those times.
- Make your coursework work for you.

    - Use classes to make connections with professors you might work with in the future.
    - Develop strategic note-taking practices.
    - Use classes to explore a topic you might research in the future.
    - Take classes with readings that overlap with your qualifying exam readings.
    - Take PE classes if they interest you.
    - Explore course offerings in other departments that have useful or interesting topics or that have joint appointment professors (e.g., education, business, law).
- Make meaningful connections.
    - Get to know faculty and more experienced graduate students in classes and workshops.
    - Don't be afraid to ask faculty to meet with you. Discussing an idea, inquiring about how you can get involved in their research, and getting their advice on research is not wasting their time.
- Envision your future after graduate school and take proactive steps to make it a reality.
    - Explore different career options (e.g., working at R1s, liberal arts schools, comprehensive universities, etc.; or working in industry)... Remember that most faculty only have the experience of working at an R1 institution, so seek out other resources on campus such as VPGE and VPTL classes and programs, ASGS workshops, and asking alumni.
    - Pursue teaching opportunities (e.g., summer session, honors college, replacement teaching, adjuncting) if they interest you.
    - Pursue industry research opportunities (e.g., internships at Facebook, SurveyMonkey, etc.; consulting; partnering with a company to use their data) if they interest you.
    - Get involved in Stanford's research centers (e.g., Center for Comparative Studies in Race and Ethnicity, Center on Poverty and Inequality, Clayman Institute for Gender Research, VMWare Women's Leadership Lab) if they interest you.
- Take advantage of Stanford's resources.
    - Get a variety of counseling services at Vaden's Counseling and Psychological Services.
    - Get confidential counseling related to assault or other forms of gender discrimination and sexual violence or report an assault.
    - Get writing or speaking tutoring at the Hume Center for Writing and Speaking.
    - Get feedback on your teaching and explore teaching resources through the Center for Teaching and Learning.
    - Get help with statistical software and data through Social Science Data and Software (SSDS).
    - Apply for funding opportunities and fellowships for Stanford graduate students.
- Be professional.
    - Have fun but remember that this is a job.
    - Use colloquia and workshops as professionalization opportunities (e.g., learning how to ask a question or give feedback).
- Use a citation software, like Mendeley, EndNote, or Zotero. Your future self will thank you!