Intro to Stata

Clay Ford, StatLab

September 10, 2014

UNIVERSITY OF VIRGINIA LIBRARY

About this workshop

Assumptions

- No prior knowledge of Stata
- Access to Stata

Teach you how to…

- Load data into Stata
- Do basic data manipulation and statistical analyses
- Save Stata syntax for reproducibility

About Stata

- Complete, integrated statistical software package
- Has both a point-and-click interface and command syntax (aka, its own programming language)
- Comes with fantastic documentation
- For Windows, Mac, and Linux/Unix
- Can purchase perpetual license or annual license
- Can purchase different flavors:
  - Stata/MP: The fastest version of Stata (for dual-core and multicore/multiprocessor computers)
  - Stata/SE: Stata for large datasets
  - Stata/IC: Stata for moderate-sized datasets
  - Small Stata: A version of Stata that handles small datasets (for students only)
- See the full sales pitch: http://www.stata.com/why-use-stata/

Getting Stata at UVa

Three ways:

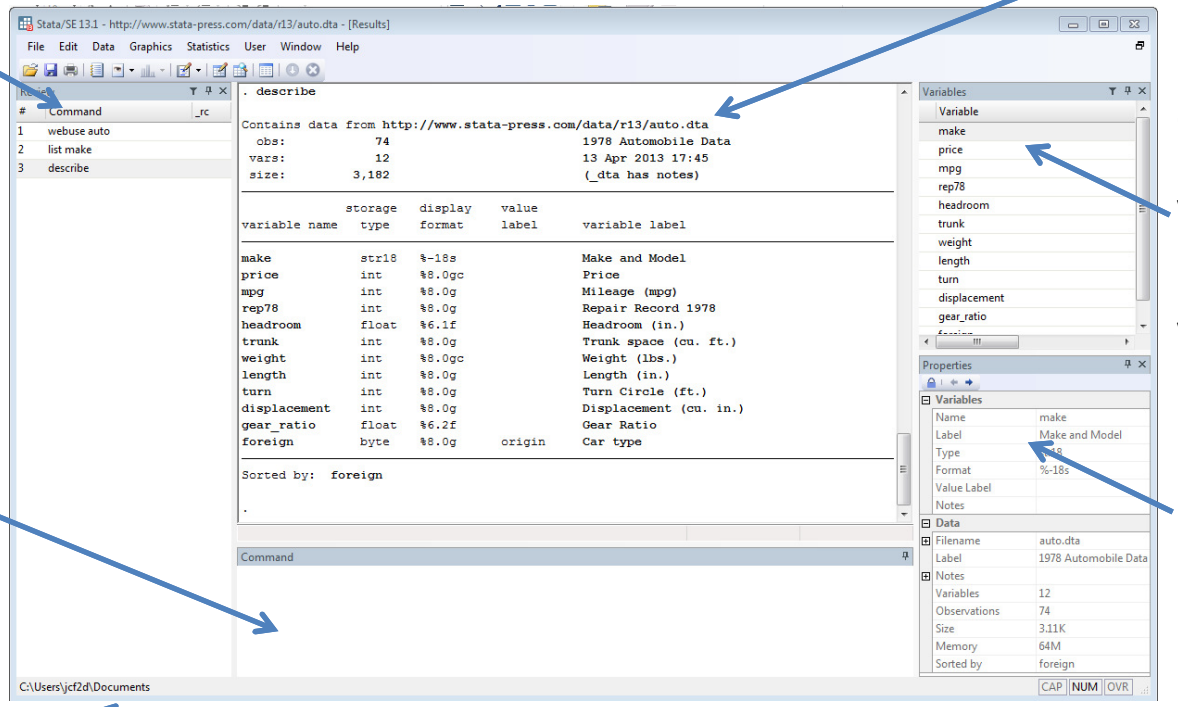
1. **Stata GradPlan**. Under this program Stata software is purchased (or rented) online and ships directly to you. See http://www.stata.com/order/new/edu/gradplans/
2. **The Hive**. Provides access to Stata/IC, version 13.
3. **Computer labs**: ITS computer lab in Clark Hall and the Scholars' Lab in Alderman Library.

Note: The University owns 45 concurrent licenses. These apply to both computer labs and the Hive. You may have to wait for licenses to open if all are in use.

Navigating Stata

Review: List of commands you have used

Command: Where you type in Stata syntax/code



Working Directory

Results: Output resulting from commands

Variables: List of variables in dataset

Properties: Information about your variables and dataset

Using Stata syntax

- Stata provides point-and-click interface but highly recommended to use syntax. Why?
  - Syntax can be saved and reused, making it easier to reproduce results and collaborate with others
  - When you Google for Stata help, the results will almost always be syntax
  - The Stata help manuals present all examples using syntax
- Syntax can be entered interactively or submitted in batch from a "do" file (ie, a text file of Stata commands)

Example of Stata syntax (set working directory and load CSV file):

```
cd "C:\Users\jcf2d\IntroToStata\"
import delimited "scores.csv"
```

Stata syntax tips

- Stata is case-sensitive
- Commands are lower case
- Most Stata commands have the following structure:
  `command variables, options`
- Many commands can be abbreviated (example, “reg” for “regress”, “gen” for “generate”, “sum” for “summary”)
- In interactive mode, previous commands can be recalled with PageUp and PageDown buttons
- Submit `help command` to get help on a command
- Submit `search topic` to find commands

Do files and Log files

- **Do** file: a text file of Stata commands; today we will work from a do file prepared in advance
  - In the do file, highlight line(s) and hit Ctrl + D to submit commands, or click the Execute Selection button: 
  - Use * or // to add comments
  - Use /// to break commands across lines
  - Use the `do` command to execute entire do file without opening
- **Log** file: record of results; captures text printed in the Results window
- When doing sustained work in Stata, it is recommended to record commands in a do file and save output to a log file

Getting started

When starting analysis in Stata:

1. Start a do file: `doedit` (or click New Do-file Editor button)
2. Set working directory: `cd "<path>"` (`"C:/Users/jcf2d"`)
   (or File…Change Working Directory)
3. Start a log: `log using "filename"` (or File…Log…Begin)

Proceed to type, submit and save commands from do file.

You can also type commands in the Command window if you're not concerned about saving commands for future use.

Let's go to the do file

I have written a do file for today's workshop. This allows you to reproduce what I have already done in Stata.

To follow along, simply highlight commands as we get to them and hit Ctrl + D (or click the submit button)

Feel free to add comments to the do file by preceding the text with an * or //

Let's go to Stata!

Getting Help/Going further

Web sites

Resources to help you learn and use Stata (IDRE UCLA)
http://www.ats.ucla.edu/stat/stata/

Germán Rodríguez's Stata Tutorial
http://data.princeton.edu/stata/

Statlist: The Stata Forum
http://www.statalist.org/

Stata NetCourses: web-based courses for learning Stata
http://www.stata.com/netcourse/

SSCC Statistical Computing Articles
http://www.ssc.wisc.edu/sscc/pubs/stat.htm

Getting Help/Going further

Books

Hamilton, L. 2012. *Statistics with STATA: Version 12*. Cengage Learning

Mitchell, M. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

Long, J.Scott. 2009. *The Workflow of Data Analysis Using Stata*. Stata Press.

Acock, Alan C. 2012. *A Gentle Introduction to Stata*, Revised 3rd ed. Stata Press.

Mitchell, Michael N. 2012. *A Visual Guide to Stata Graphics*. Stata Press.

Useful commands

General

`help` or `search` : online help on a specific command
`ssc` : access routines from the SSC Archive
`log` : log output to an external file
`tsset` : define the time indicator for time series or panel data
`compress` : economize on space used by variables
`cd` : change the working directory
`clear` : clear memory
`quietly` : do not show the results of a command

Data manipulation

`generate` : create a new variable
`replace` : modify an existing variable
`rename` : rename variable
`sort` : change the sort order of the dataset
`recode` : recode categorical variable
`drop` : drop certain variables and/or observations
`keep` : keep only certain variables and/or observations
`encode` : generate numeric variable from categorical variable
`destring` : convert string variables to numeric
`describe` : describe a data set or current contents of memory

Useful commands

More data manipulation

`use` : load a Stata data set
`import delimited` : load a text file in tab- or comma-delimited format
`save` : write the contents of memory to a Stata data set
`export delimited` : write a text file in tab- or comma-delimited format
`append` : combine datasets by stacking
`merge` : merge datasets (one-to-one or match merge)
`contract` : make a dataset of frequencies
`collapse` : make a dataset of summary statistics

Statistical commands

`tab` : abbreviation for tabulate: 1- and 2-way tables
`table` : tables of summary statistics
`summarize` : descriptive statistics
`correlate` : correlation matrices
`ttest` : perform 1-, 2-sample and paired t-tests
`anova` : 1-, 2-, n-way analysis of variance
`regress` : least squares regression
`logit, logistic` : logit model, logistic regression
`probit` : binomial probit model
`predict` : generate fitted values, residuals, etc.
`test` : test linear hypotheses on parameters

Useful commands

More statistical commands

`ivregress` : instrumental variables regression
`prais` : regression with AR(1) errors
`ologit, oprobit` : ordered logit and probit models
`mlogit` : multinomial logit model
`poisson` : Poisson regression
`heckman` : selection model
`arima` : BoxJenkins models, regressions with ARMA errors
`xtreg, (fe, re)` : fixed or random effects estimator
`xtlogit` : panel-data logit models
`xtmixed` : linear mixed (multi-level) models

Graphical commands

`histogram x:` histogram of the x variable
`twoway scatter y x:` a Y vs X scatterplot
`twoway line y x:` a Y vs X line plot
`tsline Y time:` a Y vs time time-series plot
`twoway area y x:` a Y vs X area plot
`twoway rline y x:` a Y vs X range plot (hi-lo) with lines
`twoway lfit y x:` a Y vs X least-squares fit line
`twoway lfitci y x:` a Y vs X least-squares fit line with confidence intervals
`twoway lowess y x:` a Y vs X lowess (locally weighted smoothed) line

StatLab

Thanks for coming today!

For help and advice with your data analysis, contact the StatLab to set up an appointment:
statlab@virginia.edu

Sign up for more workshops or see past workshops:
http://data.library.virginia.edu/statlab/

Register for the Research Data Services newsletter to stay up-to-date on StatLab events and resources:
http://data.library.virginia.edu/newsletters/