VIDEO: INTRODUCTION TO STATA

EMBA Data Analysis

Professor Timothy Simcoe
Boston University School of Management
Agenda

Loading data
Stata’s windows
Creating a variable
Advanced stuff
Learning resources
Loading Data

Download “Brookline Home Prices” data set
- http://people.bu.edu/tsimcoe/teaching.html

Two ways to import data
- Excel Import Tool: “File > Import > Excel...”
- As Stata (.dta) data set: “File > Open...”
- Type “clear” in Command box before loading new data

Stata always stores data as single “rectangular” worksheet
- Row = Observation, Column = Variable
Stata’s Windows

Menu Bar: Single location to access all features of program

Results Window: Where Stata displays output after you tell it to do something

Variables window: Useful info about the variables in current data set

Review window: List of previous Commands, click for a “do over”

Command window: Where you type in commands (unless you prefer to use the drop down menus)
Stata also has a “spreadsheet” view

Data Editor allows editing (not recommended)
Data Browser only permits viewing data

Click here to view data in “spreadsheet” format

Note that string variables are displayed in RED, & numeric variables are displayed in BLACK
Creating a new variable

Goal: Make new variable that measures assessed value per square foot

Using Menu Bar
- Select “Data > Create or change data > Create new variable”
- Using dialog box, provide a name (“dpsf”)... and expression (“value/area”)
- *No spaces allowed in variable names*
- Can remove variables by typing “drop [variable name]” on the Command line

Using Stata’s “generate” command
- Type “generate dpsf = value / area” on Command line

Three ways to reduce typing (and typos)
- Click next to variable name to paste in Command line
- Click in Review Window to paste previous commands
- Renaming variables, e.g. “rename dpsf warm-and-fuzzy”
Advanced stuff

Creating tabular summaries of the data
  - “table nbhd, contents(p25 value p75 value)

Visualizing the data
  - “histogram dpsf”

Modeling relationships in the data
  - “regress value rooms”
  - “twoway (scatter value rooms) (lfit value rooms)”
Additional Resources

The Instructor
- I am available to answer Stata questions any time
- I prefer to take these questions by email

Help Manuals
- Type “help” or “help [command name]” in Stata to see the manuals

Internet
- Google “Stata Help” for a wealth of information

SMG Tools
- I will post a Tutorial, Cheat Sheet and Tips to a “Stata Help” folder

Reference Books
- “A Gentle Introduction to Stata”