Introduction to Stata

Michele Claibourn

StatLab
Research Data Services
University of Virginia Library
data.library.virginia.edu

September 6, 2016

Getting to Know Stata

- A full-featured statistical programming language
- For Windows, Mac OS X, Linux
- Traditionally command-line driven syntax, though recent versions also provide point-and-click interface
- All versions (Stata/MP, SE, IC) provide the full set of features and commands (no add-ons to purchase)
- Broad range of statistical tools, strong data management features, and high-quality graphics
- Extensible, many user-provided programs
- Thousands of pages of incredible documentation
- See the full sales pitch:
  `http://www.stata.com/why-use-stata`

Navigating Stata

The default Stata user-interface consists of five windows

1. **Command**: where commands are entered for execution
2. **Results**: where all output/results are displayed
3. **Review**: a running list of all commands you've used in the order in which you've used them
4. **Variables**: displays a list of all variables in your dataset
5. **Properties**: displays properties of your variables and dataset, incl. variable names, labels, value labels, notes, formats, storage types

Wait...

Why should I type commands?

- Scientific research requires documented, reproducible research findings; using and saving syntax makes the research reproducible and collaboration easier;
- Helps you return to original data quickly and safely; makes it easier to fix any errors, perform alternate analyses (your future self will thank you);
- When you Google for Stata help, the results will almost always be syntax, help files present examples using syntax as well.

Wait...

Why should I type commands?

- Scientific research requires documented, reproducible research findings; using and saving syntax makes the research reproducible and collaboration easier;
- Helps you return to original data quickly and safely; makes it easier to fix any errors, perform alternate analyses (your future self will thank you);
- When you Google for Stata help, the results will almost always be syntax, help files present examples using syntax as well.

Is there anything else I should know?

- Stata is case sensitive; commands are always in lower case
- Most commands can be abbreviated
- When working via the Command window, previous commands can be recalled with the "Page Up" and "Page Down" buttons

Some Terminology

Log file

a full record of your Stata session; logs capture all the commands, output, and other text printed in the results window (excludes graphics).

Do-file

a text file of user-written Stata commands; do-files allow you to reproduce all manipulations and analyses of the data (including graphics), allowing for easy replication.

Commenting the do-file

- use an asterisk (*) at the beginning of a line (single line comment)
- use two forward slashes (//) at the end of any command line (single line comment)
- use the forward slash-asterisk combination (/* comment */) to bookend a comment (multi-line comment)

Today's Data

The Chronicle of Higher Education College Completion (`collegecompletion.chronicle.com/`).[1] Includes data for 3,800 degree-granting institutions. Some key variables:

- **chronname**: Institution name
- **level**: Level of institution (4-year, 2-year)
- **control**: Control of institution (Public, Private not-for-profit, Private for-profit)
- **grad_100_value** and **grad_150_value**: Percentage of first-time, full-time, degree-seeking undergraduates who complete a degree or certificate program within 100 or 150 percent of expected time
- **student_count**: Total number of undergraduates in 2010
- **med_sat_value**: Median estimated SAT value for incoming students
- **aid_value**: The average amount of student aid going to undergraduate recipients
- **endow_value**: End-of-year endowment value per full-time equivalent student
- **pell_value**: Percentage of undergraduates receiving Pell Grant

[1] Supported by the Bill & Melinda Gates Foundation.

To the do-file!

Where we'll learn how to

- set up a do-file,
- read in and save data,
- examine data,
- explore and subset data,
- manipulate and manage data,
- run preliminary descriptive statistics,
- and generate some basic graphs.

Learning More

- The Stata manuals rock – and they're available as PDFs, hyperlinked to the on-line help (type `help`)
- Stata FAQs: `http://www.stata.com/support/faqs/`
- Video Tutorials: `http://www.stata.com/links/video-tutorials/`
- Stata NetCourses: `http://www.stata.com/netcourse/`
- Statlist: The Stata Forum (`http://www.statalist.org/`) and Searchable Archive (`http://www.stata.com/statalist/archive/`)
- UCLA IDRE Resources: `http://www.ats.ucla.edu/stat/stata/`
- 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.