Session 1: Introduction to Stata

Manuel V. Montesinos

Computation Brush-Up Course
Competition and EPP Master Programs

Fall 2021

Contact

- **Google Classroom**
- **E-mail**: manuel.montesinos [at] barcelonagse.eu
- **Office hours**: upon request.

What Is Stata?

Stata is a multi-purpose statistical package to explore, summarize and analyze data. It is produced by StataCorp in College Station, TX.

It runs on Windows, Mac and Linux. The latest release is Stata 17. There is a high degree of backward compatibility, ensuring that code written for previous releases continues to work.

It is available in several **versions**, which differ in the number of variables that can be handled and the speed at which information is processed:

- Stata/IC (standard version).
- Stata/SE (extended version).
- Stata/MP (for multiprocessing).

What Is Stata?

Table 1: Versions of Stata

Product features	Stata/IC	Stata/SE	Stata/MP		
			2-core	4-core	6+
Maximum number of variables					
Up to 2,048 variables	✓	✓	✓	✓	✓
Up to 32,767 variables	-	✓	✓	✓	✓
Up to 120,000 variables	-	-	✓	✓	✓
Maximum number of observations					
Up to 2.14 billion	✓	✓	✓	✓	✓
Up to 20 billion	-	-	✓	✓	✓
Runs most estimation commands...					
Fast	✓	✓	✓	✓	✓
Twice as fast as Fast	-	-	✓	✓	✓
Almost four times as fast as Fast	-	-	-	✓	✓
Even faster	-	-	-	-	✓

Why Stata?

Stata is widely used for econometrics, biostatistics and social science research.

- **Advantages:**
  - Handling and manipulating large datasets (millions of observations!) of different types (cross-section, time series, panel).
  - Ever-growing capabilities to fit models and apply estimation techniques.
  - It can produce high-quality graphs in several different forms. Every aspect can be programmed and customized.
- **Disadvantages**:
  - Need to put the entire database into memory even though you might only be using a few variables at a time (might make it slow for large databases).
  - You can only use one dataset in memory at a time.
  - Constrained to specific file extensions.

Why Stata?

Table 2: Some Computer Tools for Applied Economics

Features	SPSS	SAS	Stata	JMP (SAS)	R	Python (Pandas)
Learning curve	Gradual	Pretty steep	Gradual	Gradual	Pretty steep	Steep
User interface	Point-and-click	Programming	Programming/ point-and-click	Point-and-click	Programming	Programming
Data manipulation	Strong	Very strong	Strong	Strong	Very strong	Strong
Data analysis	Very strong	Very strong	Very strong	Strong	Very strong	Strong
Graphics	Good	Good	Very good	Very good	Excellent	Good
Cost	Expensive (perpetual, cost only with new version). Student disc.	Expensive (yearly renewal) Free student version, 2014	Affordable (perpetual, cost only with new version). Student disc.	Expensive (yearly renewal) Student disc.	Open source (free)	Open source (free)
Released	1968	1972	1985	1989	1995	2008

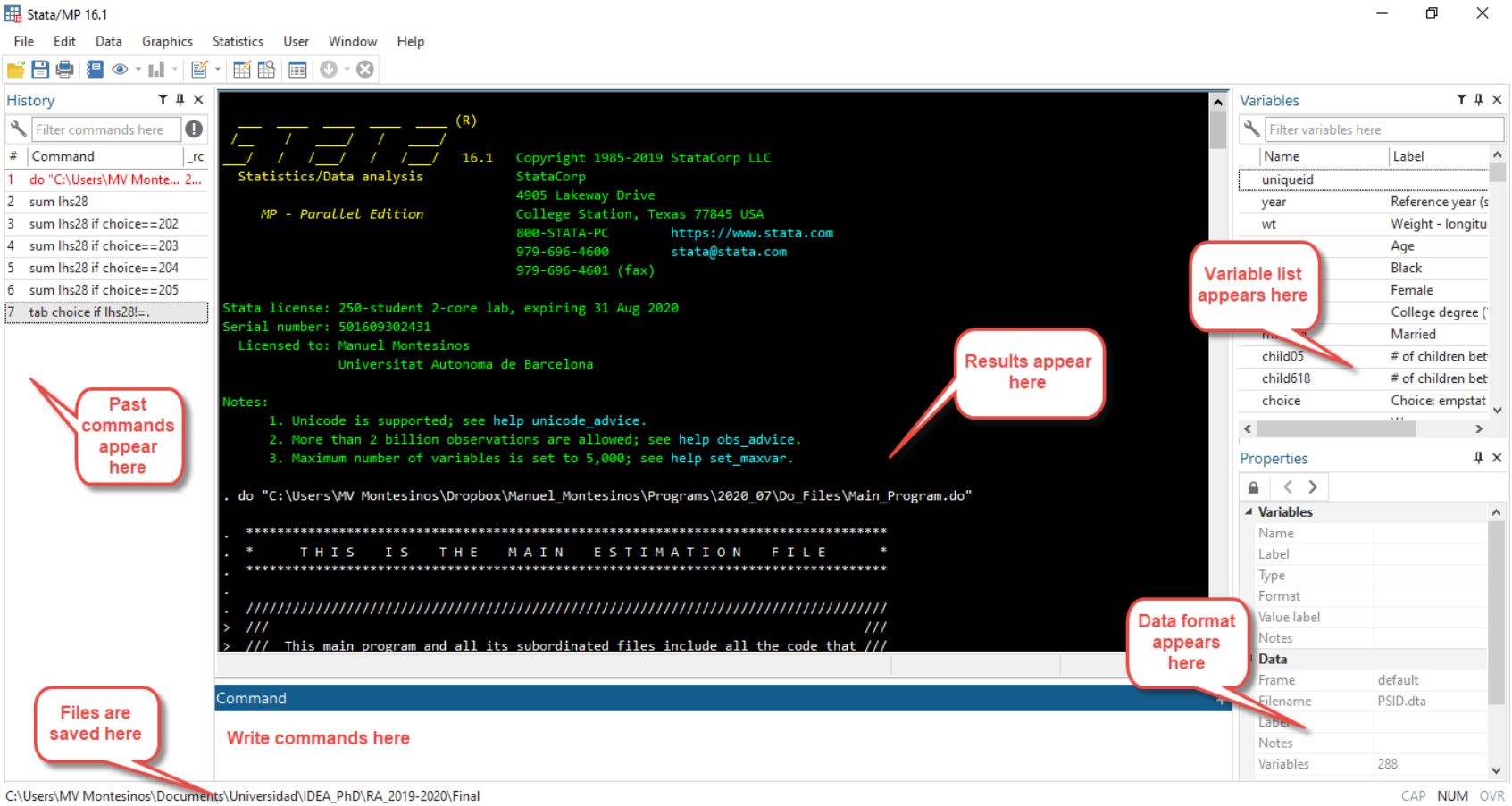
How to Get Started?

Once you have started Stata, you will see a large window containing several smaller windows:

- **Results window**: it shows operation results.
- **Variables window**: names of active variables.
- **Review window**: list of previously used Stata commands.
- **Command window**: space for typing and executing commands.

How to Get Started?

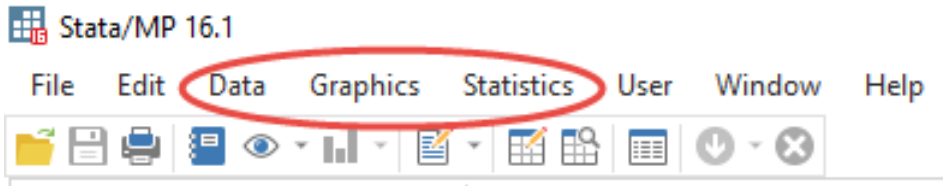
Figure 1: Stata interface



How to Get Started?

Stata is a **command-line** driven package. You can enter commands in either of 3 ways:

- **Interactively:** click through the menu on top of the screen.



- **Manually:** type the first command in the command window and execute it, then the next, and so on.

- **Do-file:** type up a list of commands in a *do-file*, an ASCII file that collects commands to be executed sequentially, and run it.

Setting Memory

RAM: Stata loads all of your data into RAM to perform its calculations. You need at least 512 MB of RAM memory for Stata to run smoothly. Make sure you have enough.

Allocation: memory allocated to Stata by the operating system while the program is running. It can be handled by the `set memory` command.

Status: to see how much memory Stata is using, and how this is being used, type `memory`.

File Extensions

Relevant file extensions:

- **.dta:** data file.
- **.do:** do-file.
- **.log:** Stata output.
- **.gph:** graphic file.

Setting Directories

Keep all the files of a specific project in **one directory**. The **goal** is to retrieve files easily and **avoid retyping long file locations**.

To see your working directory, type `pwd`. Stata will reply in the Results Window:

```
. pwd
C:/Program Files/Stata16
```

Use `dir` to check the files in your working directory.

To change the working directory, use `cd` followed by the new directory (e.g. `cd C:/mydata`).

To create a new directory, use the command `mkdir` (e.g. `mkdir C:/computation`).

Learning Stata and Looking for Help

For general help:

- **Stata online help system**: you can browse by command (typing `help command`) or by using the search box within the *Help* menu. The `findit` command looks for your command on the Internet. To learn more, type `help help`.
- StataCorp: technical support, FAQ, training courses and updates.
- StataCorp Video Tutorials.
- The Stata Blog.
- StataList – The Stata Forum.
- UCLA Stata Learning Modules.
- Oscar Torres-Reyna, "Getting Started in Data Analysis using Stata and R".

Learning Stata and Looking for Help

- Stata Course by The Econometrics Academy.

For graphs:

- Stata's visual overview for creating graphs.
- SDAS Stata Graphics Tips.

Setting up your code environment:

- Julian Reif, "Stata Coding Guide".
- Alex Hollingsworth's Workflow: Stata, GitHub, Overleaf and R.

Books:

- **Cameron, A. Colin and Pravin K. Trivedi**, "Microeconometrics Using Stata", Stata Press, Revised Edition, 2010.
- **Mitchell, Michael N.**, "Data Management Using Stata: A Practical Handbook", Stata Press, Second Edition, 2020.