

# Applied Microeconometrics

## Introduction to stata 1: Data management & descriptives

**Alessandro Martinello**

alfa 4035B

[alessandro.martinello@nek.lu.se](mailto:alessandro.martinello@nek.lu.se)



**LUND UNIVERSITY**

School of Economics and Management

# AM's reminders

- **About myself**
- **This part of the course: Hands on**
  - Quick and dirty introduction to stata
  - Short presentatuions → **You** solve problems → **I** help
- **Mail policy**
  - **NO content questions by email**
  - Ask during/between classes or just walk over to  $\alpha$  4035B

# Material

- **Uploaded before classes at Live@Lund**

- **Files** (.dta, .do) in .zip folder, **slides** uploaded separately
- **Post** .do files updated after the class
- Also see [alemartinello.com/teaching](http://alemartinello.com/teaching)

- **Structure a workspace, e.g.**

The image displays two screenshots of Windows File Explorer windows. The left window shows the 'Material' folder, which contains two subfolders: 'L1\_data\_management' (modified 13-Oct-15 8:59 AM) and 'L2\_beyond\_variables' (modified 07-Oct-15 6:48 PM). The right window shows the 'L1\_data\_management' folder, which contains five files: 'data' (folder, modified 07-Oct-15 7:01 PM), 'first\_do.do' (DO File, modified 07-Oct-15 4:40 PM), 'L1\_data\_management.do' (DO File, modified 07-Oct-15 5:01 PM), 'L1\_post.do' (DO File, modified 13-Oct-15 8:56 AM), and 'L1\_pre.do' (DO File, modified 13-Oct-15 8:59 AM).

# Why stata

- **Graphics, statistics, data management**
  - **Not GUI**  $\implies$  flexible and general
  - Does all you need
  - User written commands (e.g. **ivreg2**)
- **Easy to use & learn**
  - **Outstanding** documentation
    - UCLA, *stata* itself
    - Help files (e.g. **help regress**)
  - Intuitive syntax
- **Fast...-ish**
  - Data loaded straight into RAM (check data size)
  - If not careful, a lot of unnecessary sorting
  - “All or nothing” routines

# The workspace

The screenshot shows the Stata workspace with three main windows:

- Command window:** Contains a list of commands. The first command is `do "C:\Box Sync\Teaching\Stata_tutorial\2015\Material\Li_data_management\first_do.do"`. A blue arrow points from the text "Opens do-file editor" to the `do` command.
- Results window:** Displays the output of the command, including a table of statistics for the variable `gpm`.

Variable	Obs	Mean	Std. Dev.	Min	Max
gpm	74	.0501928	.0127986	.0243902	.0833333
- Properties window:** Shows the properties of the variable `make`, including its label "Make and Model", type "string", and format "%-18s".

Opens do-file editor

Results window

Command line

The screenshot shows the Do-file Editor window with a script named `first.do`. The script contains the following code:

```
***** A FIRST DO FILE *****/
/*
P.S.
This: /*/* means that a comment starts
possibly on more than one line.
This: /*/* ends the comment.
*/
* You can also comment on a single line after a * - only on one line!
display as text _n "Let's go!" /* Can also comment after a command with two forward slashes
*/
* You can also comment on a single line after a * - only on one line!
display as text _n "Let's go!" /* Can also comment after a command with two forward slashes
*/
let's go!
*
* you use auto, clear
(1978 Automobile Data)
*
generate gpm = 1/mpg
label var gpm "Gallons per mile"
summarize gpm
```

Do-file editor

# Running commands

- **Load data:** `sysuse`, `use` (with `using` and `[if]`)
  - e.g. `sysuse bplong`
- **Browse data:** `browse`
- **Access documentation:** `help`
- **Clear memory:** `clear`, `clear all`
- **Generate variables:** `generate`, `replace`
  - **Boolean operators:** `&` (and), `|` (or), `!` (not)
  - **Logical operators:** `==`, `!=`, `>`, `>=`, `<`, `<=`
  - **Beware missing values!**
- **Some summary stats:** `tabulate`, `summarize`, `table`

# Your turn!

- Open the example dataset **auto**
- Generate a new variable: gallons per mile
  - *Hint: variable **mpg** is miles per gallon*
- What is the % of foreign cars in the data?
- What is the average gallons per mile in the sample?
- What is the mean and max gallons per mile, by car origin (foreign/domestic)?

## Write scripts to replicate what you do

- **Material** → first\_do.do
- **CTRL + D** to execute (run does it silently)
  - Execute all or select your line
  - **iOS:** should be CMD+Shift+D(iOS 10+)



## Write scripts to replicate what you do

- **Material** → first\_do.do
- **CTRL + D** to execute (run does it silently)
  - Execute all or select your line
  - **iOS:** should be CMD+Shift+D(iOS 10+)
- Clearing the space (**clear all**) at start is often good practice

# Data management MVCs

## Handling datasets

- **drop [if]**
- **save**
- **merge**
  - Think about the data structure (panels): m:1 / 1:m / 1:1
- **append**

## Panel structure

- **bysort:** and **egen**
- To access time series shortcuts (**l.**, **d.**, **f.**): **xtset**

## Strings and labels

- **destring** and **encode**
  - **string functions:** **help functions** → String functions
- **label** (define, variable, values)