Econometric Softwares: Stata

Jérémy Do Nascimento Miguel

Class 1 Spring 2024

About myself

Jérémy Do Nascimento Miguel (check the spelling before sending an email)

- PhD candidate at Bordeaux School of Economics
- Research interests: development economics, agricultural economics, and other cool stuffs
- Office H1-B211
- Email: jdnmiguel@u-bordeaux.fr

Outline

- **Assessment**
- Syllabus
- Motivation
- Session 1
- Importing Data
- Exercise 1
- Working with data
- Exercise 2

Assessment

Two parts:

1. Submission of do-files from exercises in each class
   - → **20% of final grade**
   - → 8 do-files in total
   - → Each do-file has to be submitted until next class (so that you have enough time to practice)
2. Take-home exam
   - → **80% of final grade**
   - → Assignment with a list of tasks to be done in Stata
   - → Submission within one or two days (exact date to be decided)

Outline

- Assessment
- **Syllabus**
- Motivation
- Session 1
- Importing Data
- Exercise 1
- Working with data
- Exercise 2

Syllabus

1. Introduction to Stata and Data Management
2. Data Cleaning
3. Data Exploration and Visualizations
4. Inferential Statistics and Hypothesis Testing in Stata
5. Linear Regression and Other Modeling Techniques
6. Programming, Automation and Advanced Data Management in Stata
7. Panel Data and Survey Analysis in Stata
8. Causal Methods in Stata
9. Review, Q&A, Exam Preparation

Outline

- Assessment
- Syllabus
- **Motivation**
- Session 1
- Importing Data
- Exercise 1
- Working with data
- Exercise 2

Motivation

Why Stata?

- Fast, accurate, and easy to learn and to use
- Wide range of statistical tools in one package
- Complete control of your data and high level of flexibility in data management
- Large and continually growing number of (high quality) user contributed commands (Stata Journal)

Course Objective

- Fundamentals of data management (for empirical economic research)
- Fundamentals of programming in Stata
- Reproducible research using do-files
- Making data processing and analysis efficient
- Extracting information from data
- Implementation of fundamental regression estimations and causal analysis in Stata
- Generating clean and well-designed graphs and tables
- Advanced programming techniques

Outline

Assessment

Syllabus

Motivation

Session 1

Importing Data

Exercise 1

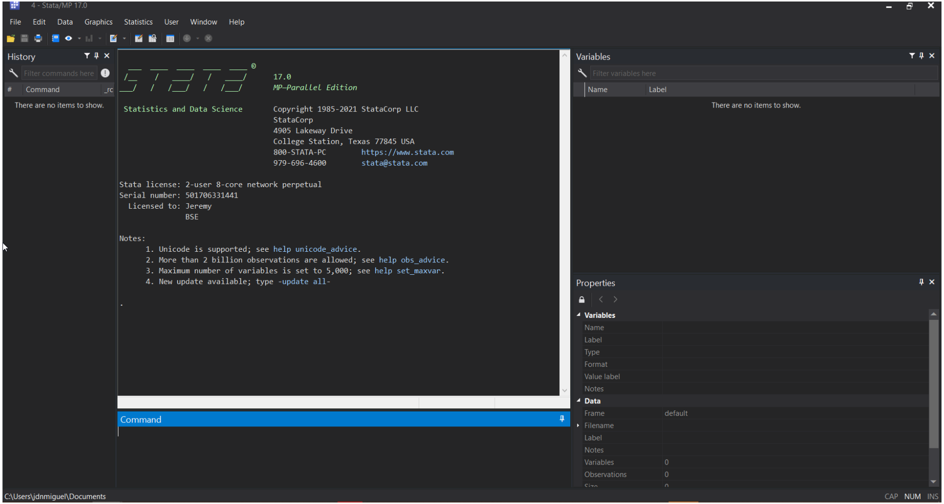
Working with data

Exercise 2

Session 1

- Getting familiar with Stata
    1. Interface, data browser
    2. Do-files
- Importing, exporting and working with data in Stata
- Basics of data processing and working with variables

Stata Interface



There are no items to show.

```
  ___  ____  ____  ____  ____ ®
 /__    /   ____/   /   ____/      17.0
___/   /   /___/   /   /___/       MP—Parallel Edition

 Statistics and Data Science       Copyright 1985-2021 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: 2-user 8-core network perpetual
Serial number: 501706331441
  Licensed to: Jeremy
               BSE

Notes:
      1. Unicode is supported; see help unicode_advice.
      2. More than 2 billion observations are allowed; see help obs_advice.
      3. Maximum number of variables is set to 5,000; see help set_maxvar.
      4. New update available; type -update all-

.
```

There are no items to show.

Variables

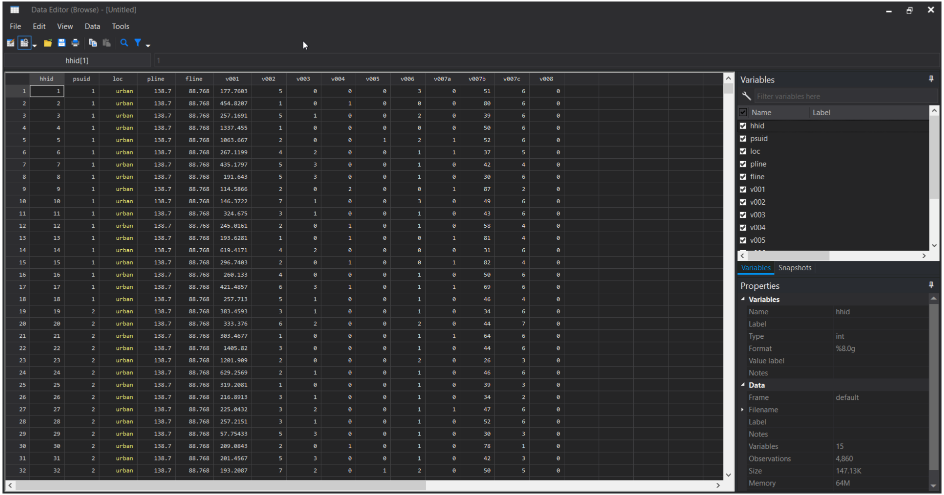
- Name
- Label
- Type
- Format
- Value label
- Notes

Data

- Frame: default
- Filename
- Label
- Notes
- Variables: 0
- Observations: 0
- Size: 0

Notes. For those interested in research, it will be your main academic friend

Browse the data



	hhid	psuid	loc	pline	fline	v001	v002	v003	v004	v005	v006	v007a	v007b	v007c	v008
1	1	1	urban	138.7	88.768	177.7603	5	0	0	0	3	0	51	6	0
2	2	1	urban	138.7	88.768	454.8207	1	0	1	0	0	0	80	6	0
3	3	1	urban	138.7	88.768	257.1691	5	1	0	0	2	0	39	6	0
4	4	1	urban	138.7	88.768	1337.455	1	0	0	0	0	0	50	6	0
5	5	1	urban	138.7	88.768	1063.667	2	0	0	1	2	1	52	6	0
6	6	1	urban	138.7	88.768	267.1199	4	2	0	0	1	1	37	5	0
7	7	1	urban	138.7	88.768	435.1797	5	3	0	0	1	0	42	4	0
8	8	1	urban	138.7	88.768	191.643	5	3	0	0	1	0	30	6	0
9	9	1	urban	138.7	88.768	114.5866	2	0	2	0	0	1	87	2	0
10	10	1	urban	138.7	88.768	146.3722	7	1	0	0	3	0	49	6	0
11	11	1	urban	138.7	88.768	324.675	3	1	0	0	1	0	43	6	0
12	12	1	urban	138.7	88.768	245.0161	2	0	1	0	1	0	58	4	0
13	13	1	urban	138.7	88.768	193.6281	1	0	1	0	0	1	81	4	0
14	14	1	urban	138.7	88.768	619.4171	4	2	0	0	0	0	31	6	0
15	15	1	urban	138.7	88.768	296.7403	2	0	1	0	0	1	82	4	0
16	16	1	urban	138.7	88.768	260.133	4	0	0	0	1	0	50	6	0
17	17	1	urban	138.7	88.768	421.4857	6	3	1	0	1	1	69	6	0
18	18	1	urban	138.7	88.768	257.713	5	1	0	0	1	0	46	4	0
19	19	2	urban	138.7	88.768	383.4593	3	1	0	0	1	0	34	6	0
20	20	2	urban	138.7	88.768	333.376	6	2	0	0	2	0	44	7	0
21	21	2	urban	138.7	88.768	303.4677	1	0	0	0	1	1	64	6	0
22	22	2	urban	138.7	88.768	1405.82	3	0	0	0	1	0	44	6	0
23	23	2	urban	138.7	88.768	1201.909	2	0	0	0	2	0	26	3	0
24	24	2	urban	138.7	88.768	629.2569	2	1	0	0	1	0	46	6	0
25	25	2	urban	138.7	88.768	319.2081	1	0	0	0	1	0	39	3	0
26	26	2	urban	138.7	88.768	216.8913	3	1	0	0	1	0	34	2	0
27	27	2	urban	138.7	88.768	225.0432	3	2	0	0	1	1	47	6	0
28	28	2	urban	138.7	88.768	257.2151	3	1	0	0	1	0	52	6	0
29	29	2	urban	138.7	88.768	57.75433	5	3	0	0	1	0	30	3	0
30	30	2	urban	138.7	88.768	209.0843	2	0	1	0	1	0	78	1	0
31	31	2	urban	138.7	88.768	201.4567	5	3	0	0	1	0	42	3	0
32	32	2	urban	138.7	88.768	193.2087	7	2	0	1	2	0	50	5	0

Name	Label
☑ hhid	
☑ psuid	
☑ loc	
☑ pline	
☑ fline	
☑ v001	
☑ v002	
☑ v003	
☑ v004	
☑ v005	

Variables	
Name	hhid
Label	
Type	int
Format	%8.0g
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	15
Observations	4,860
Size	147.13K
Memory	64M

Browse the data

Can open "Browser" window by:

- browse or br in the command window
- Clicking Data Editor (browse) button in the tool bar. Never ever you should use Data Editor (edit)

Do-file

What is it?

- a standard text file with sequence of commands to be executed in Stata
- for R users: equivalent of an R script
- must contain all steps/commands used in data preparation and analysis

You can open a do-file by:

- clicking "New Do-file Editor" button in the tool bar:
- entering doedit in the command window
- key stroke combination Ctrl+9

Do-file: Running

You can run/execute a do-file by:

- clicking “Execute (do)” button in the tool bar of Do-file Editor (whole do-file or selected lines)
- key stroke combination Ctrl+D (whole do-file or selected lines)
- typing do dofilename.do in the command window or in another do-file

The do-file should be entirely runable without any issues

Do-file: Comments

You can write down notes for your code as “comments” in a do-file:

- * : put at the beginning of a line, the line is ignored
- // : all following text/code in a line is ignored
- /* ... */ : everything in between is ignored
- /// : to split long codes into multiple lines (put at the end of a line, all following text/code in a line is ignored)

Outline

- Assessment
- Syllabus
- Motivation
- Session 1
- **Importing Data**
- Exercise 1
- Working with data
- Exercise 2

Do-file: Comments

Stata data file (*.dta)

- to open: use dataname.dta, clear

Comma-separated values file (*.csv)

- to open: import delimited using dataname.csv, clear varnames(1)

Excel sheets (*.xls or *.xlsx)

- to open: import excel using dataname.xls, clear first

clear option implies that Stata closes the data in memory to open/import selected data (!) without saving, make sure to save previous data in memory if needed!

varnames(1) or first options imply that the first row in the CSV or Excel file will be treated as variable name

Saving and exporting data

- Stata data file (*.dta): save dataname.dta, replace
- Comma-separated values file (*.csv): export delimited using dataname.csv, replace
- Excel sheets (*.xls or *.xlsx): export excel using dataname.xls, replace

replace option implies that Stata replaces the data file that already exists with the same name (if any)

Working directories and folder path

Working directory - a folder where Stata looks up files if full folder path to the file is not given in code

- Check current working directory in Stata: cd or pwd
- Changing working directory: cd your/path

Alternatively, you can directly refer to path in filenames

- But I do not recommend doing it like this

Outline

- Assessment
- Syllabus
- Motivation
- Session 1
- Importing Data
- **Exercise 1**
- Working with data
- Exercise 2

Exercise 1: 5min

1. Create a folder in your computer to store files from this session (e.g. inside the folder dedicated for this course if you have any).
2. Store the data files ”hh.csv” and ”commdist.xls” in this new folder.
3. Open a do-file and write a code which sets working directory to this newly created folder.
4. Write a code to import both data files ”hh.csv” and ”commdist.xls”.
5. Write a code to save both files as DTA files (Stata data format), with the same names (”hh.dta” and ”commdist.dta”).
6. Save your do-file in the same folder.

Outline

- Assessment
- Syllabus
- Motivation
- Session 1
- Importing Data
- Exercise 1
- **Working with data**
- Exercise 2

Documentation

To access documentation of commands and help menu:

- help command , e.g. help tabulate
- search to search for key words, e.g. search ols
- net search to search in the Internet for extra packages
- findit to search for key words locally and in the Internet

To install additional user-written commands/packages:

- ssc install command , e.g. ssc install mdesc to install mdesc command

Operators for conditions and mathematical expressions

Arithmetic operators:

- +: add
- -: subtract
- /: divide
- *: multiply
- ^: raise to the power
- log(): natural log
- exp(): exponential
- sqrt(): square root

Operators for conditions and mathematical expressions

Logical operators:

- ! or : not
- |: or
- &: and
- . or missing(): missing values

Relational operators:

- != or =:: no equal
- ==: equal
- >=: greater than or equal
- <=: less than or equal
- >: greater than
- <: less than

Variables I

Variable types:

- Numeric: continuous, integers (with or without value labels)
- String: text characters, can include anything, written inside double quotes

Generating variables:

- generate varname = exp [if]
- gen is a shortcut

Manipulating variables:

- Renaming: rename oldvarname newvarname ; but variable name can't start with number and should contain letters, numbers or underscore
- Labelling: label variable varname "label "

Variables II

- Replacing values: replace varname= exp [if] [in]
- Dropping variables: drop varlist
- Keeping variables: keep varlist
- Conditional dropping/keeping: drop x if x..

Outline

Assessment

Syllabus

Motivation

Session 1

Importing Data

Exercise 1

Working with data

Exercise 2

Exercise 2: 10 min I

1. Open the data file ”hh.dta” (use the same do-file as in Exercises 1.1.).
2. Rename following variables:
   → v001 → pccd (monthly per capita consumption); v002 → hhsize (household size) ; v003 → nchild (number of childreg aged ¡15) ; v004 → nelder (number of elderly aged 65+) ; v005 → nmigrant (number of migrants) ; v006 → nempl (number of employed) ; v007a → hhh female (HH head's gender: 1=female, 0=male) ; v007b → hhh age (HH head's age) ; v007c → hhh educ (HH head's education) ;v008 → land (area of land owned by HH)
3. Give labels to variables (variable loc defines urban/rural location; fline is food poverty line monthly, pline is total poverty line monthly, psuid is primary sampling unit ID, hhid is a unique household ID)
4. Check the age of HH head and set it to missing if age is above 90.

Exercise 2: 10 min II

5. Drop observations if households have more than six children.
6. Generate an urban dummy using variable loc.
7. Generate log of per capita consumption (add also variable label)
8. Generate employment ratio: nempl/hhsize (add also variable label)
9. Save the data with name ”hh clean.dta” (use option replace)