Stata Tutorial

Applied Microeconometrics I

Atte Pudas

Aalto University

September 2023

Introduction

- **Stata18** is available on all computers in the R102 computer lab at Väre.
- Aalto University students can get Stata18 for home use here: download.aalto.fi/student/
- Today we will cover the basics of Stata. To learn more about using Stata, see document "2023 - Stata Introduction" from MyCourses! Also, don't hesitate to google your Stata problems, the interenet is full of useful resources.

Data Set

- We will do some exercises with Finnish Longitudinal Employer-Employee Data (FLEED)
- FLEED sample data
    - created by Statistics Finland for research use
    - includes data on the person's basic characteristics, family, living, employment, relationships, periods of unemployment, income and education
- Today, we will use a small sample of it open to public.

Getting started with Stata

1. Create a new folder on your desktop.
2. Download the data set from MyCourses web page of the course and paste it into the folder you created.
3. Open Stata
   - ▶ You will immediately access the main window containing:
     - ▶ Results window;
     - ▶ Command window;
     - ▶ Review (past commands), properties (of the variables and full data set)

- ▶ Commands can be entered directly in Command window, but **do-files** are more efficient – contain the full script/Stata program.
- ▶ Strong recommendation: **always use do-files**

4. Open a new do-file: click **New Do-file Editor** on the toolbar.
5. First things to do when you start writing a script
   5.1 Change your work directory to the folder where your working files are located.

   cd "directory_name"

   5.2 Start a log-file (all output appearing in the Results window can be can be captured in a log file):

   capture log close
   log using "log_file_name.log", replace

   5.3 Load a dataset into Stata:

   use "mini_fleed_data.dta", clear

▶ To execute the code from your do-file: highlight the part of code you want to run and click Execute (do) from the toolbar or Ctrl+D keystroke

▶ You may also partially execute your code by highlighting the rows you want to execute

6. Check if you loaded your data correctly.
   ▶ by looking **Variables** window on the right top corner
   ▶ or type **browse** command in your do file and execute by highlighting it

7. Decribe your variables

describe

8. Count the number of observations

count

Note: You may always as help from STATA by using help command.

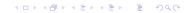**help describe**

9. Now, let's look at the content of a variable

codebook main_activity

► There are two formats of variable. "**Numerical**" vs "**String**". Mind the data format because they are handled differently in Stata.

10. Now, let's generate some new variables
    10.1 First, a continuous variable:

    gen agesq=age^2

    10.2 Next, a dummy variable:

    gen activity=0
    replace activity=1 if main_activity=="Employed" |
    main_activity=="Unemployed"

    Note: AND → & , OR → |

11. Let's check summary statistics for our generated variable

    sum activity

12. Generating more variables

   gen unemployed=.

   replace unemployed=0 if main_activity=="Employed"

   replace unemployed=1 if main_activity=="Unemployed"

13. Let's summarize unemployed

   sum unemployed

14. Now, let's look for activity and unemployment rates of women between 25 and 54 years old

   sum activity unemployed if female==1 & (age>=25 & age<=54)

15. We can also create two way tables with **table** command

   table children_under_7 female, statistic(mean unemployed) statistic(frequency)

16. We will run some OLS regressions using **regression** command. Let's regress average months of employment on number of children under 7 years old.

   16.1 First, unconditional OLS for all sample

   reg months_employment children_under_7

   16.2 Next, unconditional OLS for women and men separately

   reg months_employment children_under_7 if female==1
   reg months_employment children_under_7 if female==0

   16.3 Now, let's categorize number of children under 7

   xi: reg months_employment i.children_under_7 if female==1

- ▶ Note: We need to use "**xi: reg**" command also for string variables. Do not forget to put "**i.**" before the variables.

17. Finally, lets add some controls

    xi: reg months_employment children_under_7 i.age i.region i.education i.native_language if female==1

18. Close the log file using the command below and save the do file

    log close

19. Now, open a new do file. We will use another data set to graph some scatter plots.

20. Change your directory, open a log file and load
    "**elemapi2.dta**" into Stata

21. Draw a scatter plot of api00 and enroll variables

    twoway (scatter api00 enroll) (lfit api00 enroll)

22. Find the slope of the fitted line by regressing api00 on enroll

    reg api00 enroll

23. You can also use Stata as calculator with display command.
    Try the following examples:

    di 10*254