STATA I

Sara Nadel

sara_nadel@hksphd.harvard.edu

What is STATA???

- Computer program with capabilities to:
    - Manage (clean, organize, update) data
    - Conduct data analysis (statistics – API 209)
    - Create graphs
- Language…
    - Grammar!
    - Irregular commands!
    - Multiple ways to accomplish the same task!
    - ***Big investment now → easier to use STATA later***

Getting Started

- Copy and paste the MPAID_STATA file from the shared folder to your personal folder
- Open STATA on your computer
  - Start → All Programs → Data Analysis → STATA
- Open a *do-file*
  - Window → Do File Editor → New Do-File
  - (*also: CTRL-8)*
- Open data
  - Type the following into do file:
    cd "*path to the MPAID_STATA*"
    use maddison_gdp.dta, clear
  Do the do-file by clicking CTRL-D

What did we just do???

Windows in the main screen:

- Results: how STATA talks to us

- Command: how we talk to STATA

- Review: a list of recent commands

- Variables: list of all variables in the dataset we have open.

What is a Do-File?

- Records the commands so that you can …
  - Recreate your work if…
    - You make a mistake
    - STATA closes unexpectedly
  - Share the work we have done with others
- "do" a do-file by pressing CNTRL-D.
  - Highlight some of the commands and press CNTRL-D to do only part of a do-file.
- Make comments in your do-file by putting comments in /*between this punctuation*/

STATA Cardinal Rule #1

Always use a do-file

(except for things that are not part of the analysis – like "tasting the recipe")

Remember to save your do-file!

edit→ preferences → Auto-Save on Do/Run

The *help* command

- **he:** Help command. This explains commands and assists in identifying commands that meet your needs.
  - Test: type *help cd* into your command window

Opening and using data

- ***cd*:** change directory

- ***use*:** opens data in the STATA format
  - Type *help use* into the command window
  - *…,clear* indicates that whatever data is open should be closed before opening the new data.

Looking at data

- **count:** tells how many observations there are.
  - Type *count* into your do-file, highlight and press CTRL-D
- **describe**: gives basic information about the type of variable, and a description, if one has been added to the dataset.
  - Type *describe* into your do-file, highlight and press CTRL-D
- **summarize**: gives information about the values of a variable.
  - Type *summarize gdppc, detail* into your do-file, highlight and press CNTRL-D
- **browse**: lets us look at the data. ***Like tasting the recipe – don't put it in the do-file***
  - Type *browse* into the command window

Random STATA fact # 1

a " . " which indicates an empty value, is considered infinitely large in STATA

Using **if** in commands

- ## Operators:
  - ❑ Is exactly: "==“
  - ❑ Is less/greater than: "<“ / ">"
  - ❑ Is less than or equal to / greater than or equal to: "<=“ / ">=“
- ## Examples: (do the following commands)
  - ❑ Write *count if gdppc ==.* Into your do-file, highlight and press CTRL-D
    - ■ Tells us how many of observations in the dataset are blank for gdppc.
  - ❑ Write *summarize gdppc if country == "Chile" , detail* into your do-file, highlight and press CTRL-D
    - ■ Tells us information about the age at which people over 50 started school

Using a log file

- Type *log using STATA1.log, replace* into the second line on your do-file
- Type *log off* at the end of your do-file
- Press CTRL-D to re-do the entire do-file
- Notes on log files:
  - Cannot begin a log if one is already open
    - Highlight the *log using STATA1.log, replace* command and press CTRL-D. Do this twice.
  - Do your entire do-file by keeping everything un-highlighted and pressing CNTRL-D
    - E.g. do various parts of your do-file *other than* the log on/log off parts.

Exercises: answer the following questions

- What is the medium and the mean gdp per capita for countries in the year 1997?

- How many gdp per capita observations are missing for Argentina?

- In the years covered by the dataset, what have been the smallest and the highest gdp per capita values for Burkina Faso?

Please un-check the auto-save on run/do box before you leave for the next class.

(on do-file page, edit$\rightarrow$ preferences$\rightarrow$)

Summary

- Do-file

- Commands: help, cd, count, describe, summarize, browse, if

- Log-file

- Cardinal Rule # 1: Always use a do-file

- Random fact # 1: empty values are considered infinitely large