



Instructor: Cathy Zimmer
962-0516, cathy_zimmer@unc.edu

- 1) REMINDER – BRING FLASH DRIVES!
- 2) QUESTIONS ON EXERCISES?
- 3) WHAT IS Stata SYNTAX?
 - a) A set of commands that are saved as an ASCII/plain text file and run together.
 - b) Advantages: once saved you can use it over and over, or modify it, or copy it to do same thing many times
- 4) WORKING WITH Stata SYNTAX -- Do-files.
 - a) Creating a Do-file.
 - i) The Do-file Editor 
Click on the Do-file editor icon.

Type Stata commands in the order you want them executed. Use like any other word processor. Save by clicking File and Save as.
 - ii) Any Word Processor
Open a blank file, type the Stata commands in the order you want them executed, save as an ASCII/text file.
 - iii) Saving the Review Window contents.
Right click within the Review Window to see what can be done. Highlight a selection or select all of the Review Window contents, then click on Send to Do-File Editor. Open the Do-File editor, edit as needed, then save.
 - b) Editing a Do-file.
 - i) Edit as you would with any word processor – copy, cut, paste, save, etc.
 - ii) Including comments – add * at the beginning of a line or put text between /* and */
 - c) Running a Do-file.
 - i) In the Do-File Editor. Click the execute (do) icon. 
 - ii) In a regular Stata session. Click on File, Do and click on a do-file.

d) Saving a Do-file.

i) Click File

Save As... or click diskette button.

ii) Type in file name, keeping ".do". Choose location for file. Click Save.

EXAMPLE:

```
use "Z:\Cathy Zimmer\STATA\auto.dta", clear
* Creating a new variable - gpm.
generate gpm=1/mpg
label var gpm "Gallons per mile"
sort foreign
histogram gpm, by(foreign)
by foreign: tabulate gpm
```

Type this into a do-file, save, and execute!

5) PUTTING TWO OR MORE Stata DATA FILES TOGETHER.

a) Appending Stata Files

i) Appending places one data set directly below the other. Matching variables will be lined up, non-matching variables will be kept separate and missing values will be generated.

ii) Open the data file that you want on top.

iii) Click Data → Combine datasets → Append datasets → Browse for other dataset → OK.
OR

Enter command **append using filename**, hit enter. *filename* is the file you want on the bottom.

iv) Browse the data to make sure it looks like it should and then save it as a new data file.

v) Append these two files. Browse them first so you know what they should look like when appended.

```
“Z:\Cathy Zimmer\STATA\auto domestic.dta” (52 cases)
AND
“Z:\Cathy Zimmer\STATA\auto foreign.dta” (22 cases)
```

b) Merging Stata Files

i) Merging places one data file next to the other by matching observations. There must be one variable, with the same name, in both files that uniquely identifies a case.

ii) Sort both data files by the unique identifying variable and resave them.

iii) Open one of the data files.

iv) Click Data → Combine datasets → Merge two datasets → One-to-one on key variables → Choose key variable → Browse for other dataset → OK.

OR

Enter command **merge 1:1 varlist using filename**, hit enter. *filename* is the second data file and *varlist* is the unique identifier. – you can create one by using the following command **gen id=_n**.

v) Browse the data to make sure it looks like it should and then save it as a new data file.

- vi) Merge these two files. Browse them first so you know what they should look like when merged.

“Z:\Cathy Zimmer\STATA\auto part1.dta”

AND

“Z:\Cathy Zimmer\STATA\auto part2.dta”

6) RESHAPING Stata DATA FILES.

- a) From single observations to multiple observations per case (wide to long)
 - i) The data file must have an id variable – you can create one by using the following command **gen id=_n**.
 - ii) Look at the example data in the file I have saved as “Z:\Cathy Zimmer\STATA\wide data.dta”.
 - iii) Enter command **reshape long inc ue, i(id) j(year)**, hit enter.
- b) From multiple observations to single observations per case (long to wide)
 - i) Look at the example data in the file I have saved as “Z:\Cathy Zimmer\STATA\long data.dta”.
 - ii) Enter command **reshape wide inc ue, i(id) j(year)**, hit enter.

7) ENTERING YOUR OWN DATA INTO Stata.

- a) Make sure you have no other data files open in Stata by typing **clear all**.
- b) To open the data editor, type **edit** in the command window or click on the edit icon.
- c) Entering the Data:

Data can be entered variable-by-variable or observation-by-observation. Columns correspond to variables and rows to observations.
- d) When entering data observation-by-observation, press Tab after each value.
- e) The Tab key is smart.

After the first observation has been entered, STATA knows how many variables you have. So, at the end of the second observation (and all subsequent observations), Tab will automatically take you back to the first column.
- f) When entering data variable-by-variable, press Enter after each value.

To enter data variable-by-variable, click on the top cell in the first empty column. Type the values for the variable and press Enter after each one.
- g) Let's try it ...

You have five variables, id number, sex (m, f), race (w, b, o), years of education, job satisfaction (1 – not at all, 2 – somewhat, 3 – satisfied, 4 – very satisfied, 9 – no answer), for 10 people. Let's enter the data, name and define the variables.

Type in these values, put or leave a . in place of blanks to assign them as missing data:

```
1  m w 13 3
2  f w 16 4
3  m b . 2
4  f w 14 3
5  m w 12 9
6  f o 12 2
7  m w 12 3
8  f w 16 4
9  m w 16 1
10 f b 16 4
```

Now we need to name the variables in more meaningful ways. Click on the Variables Manager to name and label these variables. Make YOUR data file understandable to YOU!

Controlling storage using variable types:

```
byte (-127 / 126)
int (-32,768 / 32,766)
long (-2,147,483,648 / 2,147,483,646)
float (+- 10^-37 / +- 10^37)
double (+- 10^-99 / +- 10^99)
string (80 characters)
```

To store your data most efficiently, use the **compress** command at any time.

At any point, you can save as, and you will want to do so with large sets of data that you may enter, then save again and replace with the newer version. Browse your data to see how it has changed. Also try a describe command.

8) ENTERING DATA OF ANOTHER FORM INTO Stata.

a) Excel or Other Spreadsheet Files

- i) Make sure you have no other data files open in Stata by typing **clear all**.
- ii) Click File, Import, Excel spreadsheet, then fill in the needed information.
- iii) Now you can add value labels. Remember to save.

b) Delimited Text Files -- tabs or commas.

- i) Make sure you have no other data files open in Stata by typing **clear all**.
- ii) Use the command **insheet using filename**
Try it with a file I have available saved as
"Z:\Cathy Zimmer\STATA\sample data with tabs.txt"
- iii) You can do this with the menu system as well.
Click File, Import, Text data (delimited, *.csv, ...), then fill in the needed information.
- iv) Now you can add value labels. Remember to save.

- c) Blank-Delimited Text Files.
 - i) Make sure you have no other data files open by typing **clear all**.
 - ii) Open do file: "Z:\Cathy Zimmer\STATA \bball input example.do"
 - iii) Handling of character data: single quotes are not permitted.
 - iv) Handling of missing values: Use the "."

v) Input Command with Data.

```
input str12 (name team) byte points int minutes float test1
```

```
Capel UNC-CH 8 42 123.24
"Lang" "UNC-CH" 4 18 .666
Haywood UNC-CH 8 32 5.98765
Cota UNC-CH 5 40 .000008
Forte UNC-CH 28 39 666666666.2
Owens UNC-CH 0 1 454.454
Peppers UNC-CH 6 30 676767676
Coley Tulsa 6 24 .
Hill "Tulsa" 11 30 .
Kurtz Tulsa 11 26 0
Harrington Tulsa 2 33 9
Heard Tulsa 8 33 35.292929292
Davis Tulsa 0 6 .
"Swanson" Tulsa 15 18 .
Shelton Tulsa 2 17 0
Johnson Tulsa 0 8 1.111111
McDaniel Tulsa 0 5 .
end
```

```
list name team points minutes test1
```

vi) Infile Command using an external data file (containing no commands or end).

```
infile str12 (name team) byte points int minutes float test1 using "Z:\Cathy Zimmer\STATA\sample data with blanks.txt"
```

9) EXPORTING DATA FROM Stata.

- a) If you need to export Stata data so that is available for another program, you can write the STATA data as a new text file using the 'outfile' command.
- b) The command is **outfile varlist using new filename**
- c) For example, type **outfile team points using "H:\bball.txt"** – use the drive you need.

EXERCISES:

Using data from the 2018 General Social Survey, create and run a do-file to ...

1. Open the data set (Z:\Cathy Zimmer\STATA \2018 u.s. general social survey.dta).
2. How many cases are there? How many variables?
3. Produce a frequency distribution for the **happy** variable for females only. How many cases are missing data?
4. Calculate a new variable that is the average of mother's and father's education. Call it **avged**.
5. Produce a histogram for the new variable. How many cases are missing data?
6. Run a frequency distribution for number of siblings (**sibs**). Then recode into a new variable called **catsibs** with FOUR categories: none, one, two, three or more. Check your recoding by running a frequency distribution and comparing it to the original one.