

Unit 4– Introduction to Stata *version 16*

Homework 1 of 2

Due: Wednesday September 23, 2020

Last date for submission with -10 points: Friday September 25, 2020

Last date for submission for credit (-20 points): Wednesday September 30, 2020

Introduction

This homework gives you practice with some of the basics.

Preliminary

Download from the course website to your desktop, the Stata data set [wws1000.dta](#).

You can find it here

<http://people.umass.edu/biostat690c/webpages/stataintro.html>

Preliminary

Launch Stata and begin a log of your session. Take care to save it as a “.log” and not as a “.smcl” file.

Preliminary

Best Practices: In the command window, type in the following preliminary commands, editing **mine** as needed to become yours!

```
* BIOSTATS 690c – Fall 2020
*      prog:  Carol Bigelow
*      date:  9-16-2020
*      input:  wws1000.dta
*      output: none
*      title:  Introduction to Stata (1 of 2)
```

Exercises

- ___1. **This exercise gives you practice in learning about the structure of a data set.**
 - a) Execute the command **describe**.
 - b) Execute the command **codebook, compact**.

- ___2. **This exercise gives you practice in listing data set.**
 - a) Type the command **help list**. Scroll down to learn how to print first and last observations.
 - b) Print the first 5 observations of *race*, *age*, *married* and *wage*.

- ___3. Print the last 5 observations of *south*, *industry* and *occupation*.

- ___4. **This exercise gives you practice with tab2 for two way cross tabulations.**
 - a) Type the command **help tab2**. Scroll down to learn about the various options.
 - b) Produce a two-way cross-tabulation of *race* and *married* with row percentages.
 - c) Produce a two-way cross-tabulation of *race* and *married* with column percentages
 - d) Produce a two-way cross-tabulation of *race* and *married* with cell percentages

- ___5. **This exercise gives you practice with the command sort. It also gives you an introduction to the command table for obtaining summary statistics for a continuous variable, for each level of a discrete variable.**
 - a) Type the command **help sort**. Scroll down to learn!
 - b) Type the command **help table**. Scroll down to learn!
 - c) Execute the command **table** with the options **contents** and **by()** to obtain the following statistics for the variable *wage*, separately for each level of the “by” variable *race*:: mean, standard deviation, minimum and maximum

- ___6. **This exercise gives you practice with command sample to sample from a data set.**
 - a) Re-read page 58 of the Unit 4 lecture notes.
 - b) Type the command **help sample**. Scroll down to learn about the various options.
 - c) Obtain a sample of size 10.
 - d) Save this to a new data set called *wws_sample10.dta*.

- ___7. **This exercise gives you practice exporting a STATA data set to EXCEL.**
 - a) Reread page 33 of the Unit 4 lecture notes.
 - b) Export *wws_sample10.dta* to an EXCEL data set that you save as *wws_sample10.xls*

- ___8. **Close your log**

- 9. **ALWAYS produce a clean and error-free record of your work.**
- a) Launch Word and open your saved log.
 - b) Edit so as to eliminate all your mistakes
 - c) Save to a PDF
 - d) Upload to the Blackboard Learn ASSIGNMENT tab