

Unit 4– Introduction to Stata *version 16*

Homework 2 of 2

Due: Thursday October 8, 2020

Last date for submission for credit (-10 points): Thursday October 15, 2020

Introduction

This homework gives you practice with data cleaning and the creation of new variables. You will also gain practice in working with DO files (handy....)

Preliminaries

- __ a) Be sure that you have already downloaded from the course website the data set *introstata1.dta*
- __ b) Be sure that you have already downloaded from the course website the data set *introstata2.dta*
- __ c) Launch MS Word. Begin a word document that will house your work.

Preliminaries – Chart of Possible Data Screens

	Screen	Stata Commands
1	Review data and codes	. describe . codebook . codebook, compact . label list
2	List data values <i>Be careful with this!</i>	. list
3	Identify duplicates	. duplicates
4	Identify missing values	. list
5	Range check	. tabstat
6	Identify out of range values	
7	Screen for nominal variable	. tabulate
8	Screen for continuous variable	. inspect
9	Screen for inconsistency	. tab2 . list
10	Detailed summary for discrete	. tabulate
11	Detailed summary for continuous variable	. summarize

EXERCISES #1 – 7

introstata1.dta

Data Dictionary

Variable	Variable Label	Type	Coding
studyid	Unique study id	Character	
city	Does woman live in metro area?	Numeric	1=yes, 0=no
dues	Union dues paid last week	Numeric, continuous	

1. Before launching STATA.

This exercise gives you practice in planning your data screens. Using a design of your choosing, create a table that specifies the particular data checks/cleans that you will do for each of the three variables in this data set (*Tip* – Use the chart on the previous page as a guide).

Launch stata and begin a log of your session. Take care to choose the “.log” extension.
Open the data set *introstata1.dta*

2. This exercise gives you practice in defining variable labels.

3. This exercise gives you practice in defining variable value code labels.

4. This exercise gives you practice in data screening.

Using the plan you developed in exercise #1, perform the data screens you designed. ***Tip!* Repeat each of your commands as many times as necessary to obtain an execution that is correct. This is necessary in order to end up with a do-file that will execute without error.**

5. This exercise gives you practice in saving your data set.

Save your changes to the data set into a new Stata data set that you name *mystatadata1.dta*.

6. This exercise gives you practice in creating a do-file.

- Position your cursor (mouse) in the REVIEW window. RIGHT CLICK
- From the drop down menu, choose SAVE ALL
- Save your commands to a file that you name *mydofile1.do*
***Tip!* Be sure that you provide the extension “.do” if it is not provided for you.**

7. This exercise gives you practice in finalizing a do-file.

- By any means you like (in Stata or a text-editor) open your do file.
- Edit out ALL the mistakes
- Save the cleaned version.

EXERCISES #8 – 13*introstata2.dta*

Tip! Always familiarize yourself with the data. In Stata, after you have opened *introstata2.dta*, issue the command **codebook, compact**. Alternatively, issue the command **describe**.

___ 8. **This exercise gives you practice creating new variables using mathematical functions available in Stata.** **Tip –** To get you on your way, issue the command **help functions** before doing this exercise. Create a new variable **lnlead** that is the natural logarithm of the variable **lead**. Label it.

___ 9. **This exercise gives you practice creating new variables that are a function of existing variables.** Create a new variable **bmi** using the variables **height** and **weight** and using the definition

$$\text{bmi} = \frac{\text{weight(kg)}}{[\text{height(m)}]^2} = \frac{(100)^2 * \text{weight(kg)}}{[\text{height(cm)}]^2}$$

___ 10. **This exercise gives you practice creating a new variable that is a grouping of a variable.** Create a new variable **bmi_group** using the variable **bmi** and defined as follows:

$$\begin{aligned} \text{bmi_group} = & \begin{array}{ll} 1 \text{ if } \text{bmi} \leq 18.49 & (\text{underweight}) \\ 2 \text{ if } 18.5 \leq \text{bmi} \leq 24.99 & (\text{normal}) \\ 3 \text{ if } 25.0 \leq \text{bmi} \leq 29.99 & (\text{overweight}) \\ 4 \text{ if } \text{bmi} \geq 30.00 & (\text{obese}) \end{array} \end{aligned}$$

___ 11. **This exercise gives you practice creating discrete value labels and variable name labels** Create discrete variable labels for **bmi_group**. Label the variable **bmi_group**. As a check on your work, issue the command **tabulate bmi_group**.

___ 12. **This exercise gives you practice in defining reverse coded variables.** Create a new variable called **racer** that is a reverse coding of the variable **race**. As a check on your work, issue the command **tab2 race racer**.

___ 13. **Save your data set.** Save your changes to the data set into a new Stata data set that you name *mystatadata2.dta*.

_____ **Package your work to a clean report for submission to the ASSIGNMENTS tab in Blackboard**

- (a) In Stata, close your log
- (b) Exit Stata
- (c) Make active your WORD session
- (d) Now open your stata log and paste it into your word document
- (e) **Clean it up!!!** That is – delete all the error messages and false starts, etc.
- (f) **Fix the font!!!** If your report looks all wavy and nothing lines up properly, this is because the font is not right. Select the entire document and then change the font to something that lines up properly, such as courier.

Please please please!

Do **NOT**

Submit a report that has errors in it

You wouldn't do this to a client, right?

_____ **Finally, save your work to a pdf file (print to a pdf) and submit to the ASSIGNMENTS tab in Blackboard.**