

## Unit 4– Introduction to R

### 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....)

#### Preliminary. Datasets and Word document.

- \_\_\_ a) Download from the course website the data set *intror1.Rdata*
- \_\_\_ b) Download from the course website the data set *intror2.Rdata*
- \_\_\_ c) Launch MS Word. Begin a word document that will house your work.

#### Preliminary – Packages

- \_\_\_ a) In R Studio, be sure you have already done a one-time installation of the package *Hmisc*

#### Preliminary – Initialize your session

- \_\_\_ a) Launch R Studio
- \_\_\_ b) From the tool bar at top, open a new R Markdown file; delete everything below the header
- \_\_\_ c) Save it under the appropriate name using the convention **lastname\_hw4.Rmd**
- \_\_\_ d) Using the command **setwd( )**, set the working directory to the folder containing the data
- \_\_\_ e) Load the two datasets **intror1.Rdata** and **intror2.Rdata**

#### How to approach this assignment:

In each exercise below, you are asked to create a new chunk. In it, you will type your code, send it to the console for execution. As errors occur, you then modify the code in the chunk and re-execute. You will continue cycling in this way until the chunk runs with NO error. Then save your R Markdown file and move on to the next exercise.

**EXERCISES #1 – 2**

Exercises #1-2 assume that, in your R Studio session, you have issued the command `load( )` and loaded the R dataset `intror1.Rdata`. If you now look in the ENVIRONMENT pane, you should see the object named `intror1`.

## Data Dictionary

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

- \_\_\_ 1. **This exercise gives you practice in defining variable labels.**  
In a new chunk: Define all variable labels.
- \_\_\_ 2. **This exercise gives you practice in defining variable value code labels.**  
In a new chunk: Define variable value codes for ONLY the discrete variables in this data set.  
Issue the command `table( )` to produce a table of the distribution of the discrete variable `city`

**EXERCISES #3 – 6**

Exercises #1-2 assume that, in your R Studio session, you have issued the command `load( )` and loaded the R dataset `intror2.Rdata`. If you now look in the ENVIRONMENT pane, you should see the object named `intror2`.

- \_\_\_ 3. **This exercise gives you practice creating new variables using mathematical functions available in R.**  
In a new chunk: Create a new variable `lnlead` that is the natural logarithm of the variable `lead`. Label it. Issue the command `summary( )` to produce summary statistics for the continuous variable `lnlead`
- \_\_\_ 4. **This exercise gives you practice creating new variables that are a function of existing variables.**  
In a new chunk: Create a new variable `bmi` using the variables `height` and `weight` and using the definition on the next page.

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

Issue the command **summary( )** to produce summary statistics for the continuous variable **bmi**

\_\_\_5. **This exercise gives you practice creating a new variable that is a grouping of a variable.**

In a new chunk: Create a new variable **bmi\_group** using the variable **bmi** and defined as follows:

<b>bmi_group</b> = 1 if $\text{bmi} \leq 18.49$	(underweight)
2 if $18.5 \leq \text{bmi} \leq 24.99$	(normal)
3 if $25.0 \leq \text{bmi} \leq 29.99$	(overweight)
4 if $\text{bmi} \geq 30.00$	(obese)

Issue the command **table( )** to produce a table of the distribution of **bmi\_group**

\_\_\_6 **This exercise gives you practice creating discrete value labels and variable name labels**

In a new chunk: Create discrete variable labels for **bmi\_group**. Label the variable **bmi\_group**. Again, issue the command **table( )** to produce a table of the distribution of **bmi\_group**

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

- Knit your “error free” R Markdown file to MS WORD
- Then launch MS WORD – make any additional edits you like
- Save (print to) a PDF
- Upload to Blackboard Learn