ECON 203 – Economic Statistics II
Department of Economics – University of Illinois at Urbana-Champaign
Fifth Assignment
Delivery Date: March 4, 2024, 11:59 pm

The fifth assignment consists of two questions. The first is empirical and the second question is related to the empirical part and should be answered directly on Canvas.

Your answers to the homework assignments must be completed **individually**.

The following rules apply:

- The practical questions involving programming should be delivered as R codes (.R file) and a PDF file containing the outputs of the code (tables, plots, etc).
- The answers must be uploaded on Canvas by the due date and time. Late homework will not be accepted. Please use the following convention to name your files: **_HW[number]_LastName_FirstName**.

Question	Points	Bonus Points	Score
1	20	0	
2	80	0	
Total:	100	0	

No not write on the table above.

Good Luck!

1. In this assignment, you will continue to work with the `housing.xls` file. The dataset was collected from the real estate pages of the Boston Globe in 1990. These homes were sold in the Boston, MA area. There are 88 observations in the dataset and the following variables:

price	house selling price, measured in \$1000s
assess	assessed value, measured \$1000s (value before the house was sold)
bdrms	number of bedrooms
lotsize	size of lot in square feet
sqrft	size of house in square feet
colonial	= 1 if home is colonial style or = 0, otherwise

The first question consists of loading the dataset in R Studio and running some basic analysis. To load the data, you should follow the steps below:

1. Open R Studio on your computer;
2. install the package `readxl`. To install a package, you should use the function `install.packages`;
3. load the `readxl` library. You should use the function `library`;
4. define the location of the `housing.xls` file on your computer. Use the function `setwd`;
5. load the data with the function `read_excel`.

You can check if the data have been correctly loaded using the function `head`. Figure 1 shows how the code will look after following the above instructions.

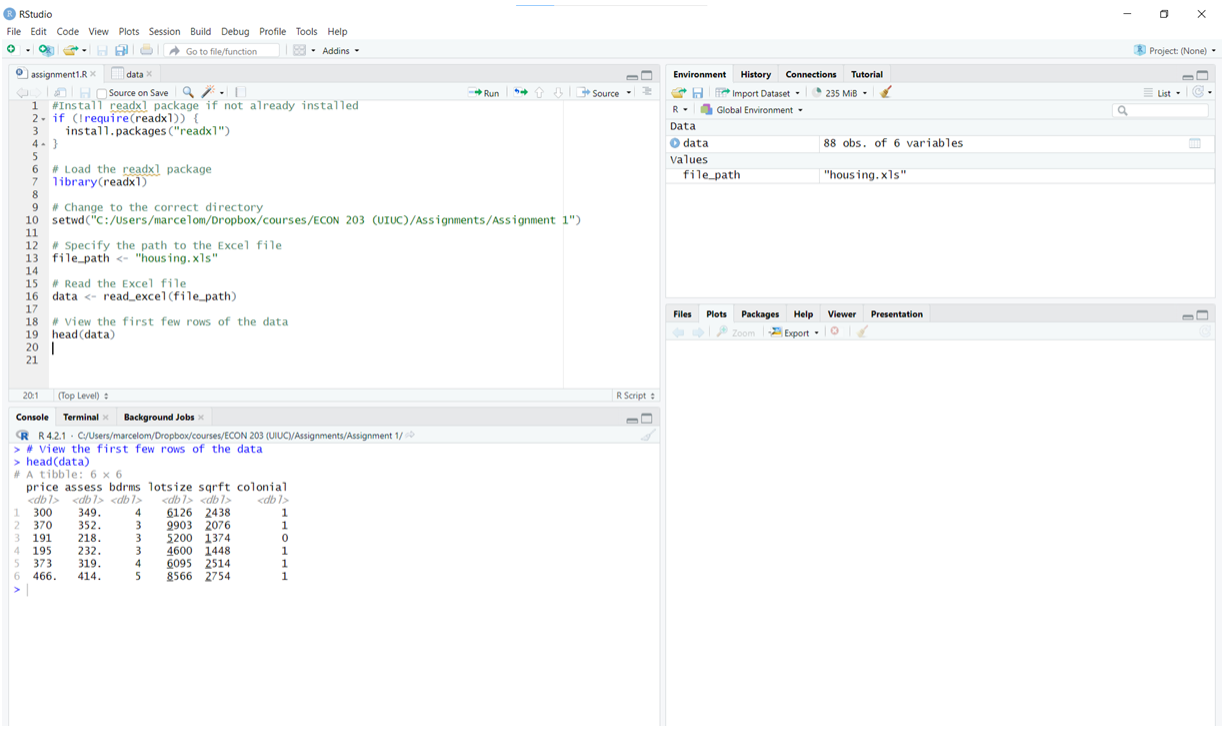

Figure 1: R studio output

Suppose that you own a real state business in the Boston area and imagine that we are back in the 1990s. Based on the dataset you have, your goal is to understand the behavior of the market.

(a) (3 points) Compute the average price for colonial and non-colonial houses.

(b) (5 points) Now, you will estimate a linear model relating house prices with the style of the house (colonial or non-colonial). Therefore, you want to estimate:

$$\mathsf{Price}_i = b_0 + b_1 \mathsf{Colonial}_i + U_i,$$

where U_i is the error of the model.

Remember that we learned in class that the optimal estimators for b_0 and b_1 are given by:

$$\widehat{b}_0 = \overline{\mathsf{Price}} - \widehat{b}_1 \overline{\mathsf{Colonial}}$$

$$\widehat{b}_1 = \frac{\sum_{i=1}^{n} \left(\mathsf{Price}_i - \overline{\mathsf{Price}}\right)\left(\mathsf{Colonial}_i - \overline{\mathsf{Colonial}}\right)}{\sum_{i=1}^{n} \left(\mathsf{Colonial}_i - \overline{\mathsf{Colonial}}\right)^2}$$

You can compute $\widehat{b}_0$ and $\widehat{b}_1$ using the functions `sum` and `mean`. Compare your results with the ones from the previous question. What is the interpretation for b_0 and b_1 in the model above?

(c) (2 points) Now you will learn an easier way to estimate a linear model in R with the function `lm`. To estimate the previous model simply type `linear_model <- lm(data$price ~ data$colonial)` (assuming that you data variable is called `data`).

You can access the estimates by typing `linear_model$coefficients`.

(d) (5 points) Repeat the previous item replacing Colonial by sqrft. Draw a scatter plot of Price against sqrft. Draw on top of the scatter plot the best-fitting line. Use the function `abline`.

(e) (5 points) Repeat the previous item replacing sqrft by lotsize. Draw a scatter plot of Price against lotsize. Draw on top of the scatter plot the best-fitting line. `abline`.

2. This question must be answered on Canvas.

(a) (10 points) The average price for colonial houses is _____. Answer with two decimal digits

(b) (10 points) The average price for non-colonial houses is _____. Answer with two decimal digits

(c) (10 points) The estimated b_0 and b_1 on item (b) of Question 1 are _____ and _____. Answer with two decimal digits

(d) (10 points) The value of b_0 in the equation of Question 1 (b) is the average price of non-colonial houses. True or False?

(e) (10 points) The value of b_1 in the equation of Question 1 (b) is the average price of colonial houses. True or False?

(f) (10 points) The estimated b_0 and b_1 on item (d) of Question 1 are _____ and _____. Answer with two digits.

(g) (10 points) The estimated b_0 and b_1 on item (e) of Question 1 are _____ and _____. Answer with two decimal digits.

(h) (10 points) The variable `lotsize` seems to explain more the difference in `Prices` than the variable `sqrft`. True or False?