

ECON 203 – Economic Statistics II
Department of Economics – University of Illinois at Urbana-Champaign
Second Assignment
Delivery Date: February 5, 2024, 11:59 pm

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

This assignment aims to teach how to compute simple probabilities in R using basic instructions and how to write a loop. You do not need advanced programming skills to complete this assignment.

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

The following rules apply:

- The **TRUE** answers should be completed on Canvas.
- 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	60	0	
2	40	0	
Total:	100	0	

No not write on the table above.

Good Luck!

1. In this assignment, you will 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.

The screenshot shows the R Studio interface with the following components:

- Source Editor:** Contains R code to install `readxl`, load it, set the working directory, and read the `housing.xls` file into a data frame named `data`. It also includes a `head(data)` command to view the first few rows.
- Environment Pane:** Shows the `data` object as a data frame with 88 observations and 6 variables. The `file_path` variable is also listed with the value `"housing.xls"`.
- Console:** Displays the output of the `head(data)` command, showing a tibble with 6 rows and 6 columns: `price`, `assess`, `bdrms`, `lotsize`, `sqrft`, and `colonial`.

```
1 #Install readxl package if not already installed
2 if (!require(readxl)) {
3   install.packages("readxl")
4 }
5
6 # Load the readxl package
7 library(readxl)
8
9 # Change to the correct directory
10 setwd("C:/Users/marcelom/Dropbox/courses/ECON 203 (UIUC)/Assignments/Assignment 1")
11
12 # Specify the path to the Excel file
13 file_path <- "housing.xls"
14
15 # Read the Excel file
16 data <- read_excel(file_path)
17
18 # View the first few rows of the data
19 head(data)
20
21
```

Console output:

```
> # View the first few rows of the data
> head(data)
# A tibble: 6 x 6
  price assess bdrms lotsize sqrft colonial
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 300 349. 4 6126 2438 1
2 370 352. 3 9903 2076 1
3 191 218. 3 5200 1374 0
4 195 232. 3 4600 1448 1
5 373 319. 4 6095 2514 1
6 466. 414. 5 8566 2754 1
```

Figure 1: R studio output

Before answering the questions below, read Sections 1.1-14 from the file “Heiss (Using R for Introductory Econometrics).pdf,” available from Canvas. Look at the module “Learning the R programming language.”

(a) (10 points) Define the following random variables:

- A : the number of bedrooms in a house
- P : the price of a house

Using R, compute the probability $P(P > 200|A = 3)$. Remember that

$$P(P > 200|A = 3) = \frac{P(P > 200 \cap A = 3)}{P(A = 3)}.$$

The first step is to compute $P(P > 200 \cap A = 3)$. To compute such probability, you must first select the prices for three-bedroom houses.

Now, to compute $P(P > 200 \cap A = 3)$, you should compare the proportion of three-bedroom houses with prices larger than 200. To do so, you can use the function `sum` in R. It sums the elements in a vector. For instance, to compute the number of three-bedroom houses with prices larger than 200, you can use the function below:

$$\text{sum}(\text{data}\$price[\text{data}\$bdrms == 3] > 200).$$

(b) (30 points) Now, the goal is to compute the above probability for different prices and number of rooms, i.e., we want to compute

$$P(P > p|A = a) = \frac{P(P > p \cap A = a)}{P(A = a)},$$

for $a = 2, 3, 4$ and $p \in \{100, 101, 102, 103, \dots, 499, 500\}$. You need to learn how to construct loops in R to do this easily. Please read section 4.2 (pages 42-46) in the R manual uploaded on Canvas. You can find the manual in the module labeled “Learning the R programming language”.

An example of a loop in R is as follows:

```
x <- rep(0, 10)
for(i in 1 : 10){
  x[i] <- i + 10
}
```

The code above will create a vector x with ten positions. At each position i the value $10 + i$ will be stored.

(c) (20 points) Construct plots of $P(P > p|A = a)$ against p . You should construct three plots. One for $a = 2$, another for $a = 3$, and a third one for $a = 4$.

2. The second question should be answered directly on Canvas.

- (5 points) What is the value of $P(P > 200|A = 4)$? Answer with three decimal points.
- (5 points) What is the value of $P(P > 200|A = 3)$? Answer with three decimal points.
- (5 points) What is the value of $P(P > 200|A = 2)$? Answer with three decimal points.
- (5 points) What is the value of $P(P > 300|A > 3)$? Answer with three decimal points.
- (5 points) What is the value of $P(200 < P < 300|A = 3)$? Answer with three decimal points.

- (f) (5 points) How many houses in the sample have more than three bedrooms?
- (g) (5 points) What is the average price of three-bedroom houses? Report only the integer part. Note that prices are in thousands of dollars. However, you **DO NOT** need to multiply it by 1,000. Just report the integer part of the result you have obtained in R.
- (h) (5 points) What is the average price of four-bedroom houses? Report only the integer part. Note that prices are in thousands of dollars. However, you **DO NOT** need to multiply it by 1,000. Just report the integer part of the result you have obtained in R.