Assignment 6

Question 1:
Answer the following questions using the dataset TeachingRatings.rda that you will find in the Data folder on Learn.

The data is taken from a study published in 2005 in the Economics of Education Review. The title of the paper is “Beauty in the Classroom: Instructors’ Pulchritude and Putative Pedagogical Productivity”. The goal is to measure the effect of instructor looks on the students evaluation. Beauty has been shown to be related to either wage or productivity in other areas. This study tries to see if it also applies to education.

The data are from 463 courses at the University of Texas at Austin from 2000 to 2002. The variables included in the dataset are:

- minority: Does the instructor belong to a minority (non-Caucasian) (yes or no)
- age: Professor’s age in years
- gender: A factor indicating the instructor’s gender (male or female)
- credits: Is the course a single-credit elective (single or more)
- beauty: Rating of the instructor’s physical appearance. The scale is centered to have a mean of zero. The range goes from -1.45 to 1.97.
- eval: The course evaluation (1 to 5)
- division: Is the course a lower (first or second year) or upper level (lower or upper)
- native: Is the instructor a native English speaker? (yes or no)
- tenure:. Is the instructor on tenure track? (yes or no)
- students: The number of students that participated in the evaluation.
- allstudents: The number of students enrolled in the course.

There is something new to you in this assignment. Some variables are factors with words instead of numbers. For example, the variable gender contains either the word ”male” or ”female”.

```
data$gender[1:10]

##  [1] female male   male   female female male   female female female male
## Levels: male female
```

The levels indicate what different values the variable can take. Fortunately, you do not need to worry because R known how to deal with such variables. For example, if you run a regression with the factor you get the following:

```
res <- lm(eval~gender, data=data)
summary(res)

##
## Call:
## lm(formula = eval ~ gender, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.96903 -0.36903  0.03097  0.43097  0.99897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.06903    0.03355  121.29  < 2e-16 ***
## genderfemale -0.16800    0.05169   -3.25  0.00124 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5492 on 461 degrees of freedom
## Multiple R-squared:  0.0224,Adjusted R-squared:  0.02028
## F-statistic: 10.56 on 1 and 461 DF,  p-value: 0.001239
```

The variable name is genderfemale, which means that R has created a variable equals to 1 if it is female and 0 otherwise and put it in the regression. You can even use the predict() function and ignore how R deals with the variable. Suppose you want to predict the average evaluation of females and males, you would do

```
predict(res, newdata=data.frame(gender="male"), interval="confidence")

##       fit      lwr      upr
## 1 4.06903 4.003103 4.134957

predict(res, newdata=data.frame(gender="female"), interval="confidence")

##        fit      lwr      upr
## 1 3.901026 3.823738 3.978314
```

For this assignment, you do not need to know how to interpret qualitative variables. We will cover that in Topic 7. You will only have to perform tasks similar to what you did in the tutorial session.

1. Estimate the following model:

$$eval = \beta_0 + \beta_1 beauty + \beta_2 I(beauty^2) + u_i$$

   Interpret your result. Should $beauty^2$ be included in the model?

2. Compute the average partial effect of beauty on the course evaluation.

3. Construct a 95% confidence interval for the average partial effect of beauty on the course evaluation. Do it manually.

4. Construct a 95% confidence interval for the average partial effect of beauty on the course evaluation. Do it using a proper transformation of the model.

5. Estimate the following model:

$$eval = \beta_0 + \beta_1 beauty + \beta_2 age + \beta_3 I(beauty * age) + u_i$$

   Interpret your results.

6. Using the new model, compute the average partial effect of beauty on the course evaluation.

7. Using the new model, construct a 95% confidence interval for the average partial effect of beauty on the course evaluation. Do it manually.

8. Using the new model, construct a 95% confidence interval for the average partial effect of beauty on the course evaluation. Do it using a proper transformation of the model.

9. Select the best model among the following ones using the adjusted R^2.

$$eval = \beta_0 + \beta_1 beauty + \beta_2 age + \beta_3 I(beauty * age) + u_i$$

$$eval = \beta_0 + \beta_1 beauty + \beta_2 age + \beta_3 I(age^2) + \beta_4 I(beauty * age) + u_i$$

$$eval = \beta_0 + \beta_1 beauty + \beta_2 age + \beta_3 I(beauty * age) + \beta_4 I(beauty^2) + u_i$$

$$eval = \beta_0 + \beta_1 beauty + \beta_2 age + \beta_3 I(beauty * age) + \beta_4 male + u_i$$

$$eval = \beta_0 + \beta_1 beauty + \beta_2 age + \beta_3 I(beauty * age) + \beta_4 male + \beta_5 native + u_i$$

   Then, estimate the best one and interpret the result. Before doing it, create the male and native variable this way:

```
data$male <- as.numeric(data$gender == "male")
data$native <- as.numeric(data$native == "yes")
```

You do not need that for the regression, but it will be easier to transform the model in the other questions

10. Using the best model, construct the 95% confidence interval for the average evaluation of 40 year old male professors with a beauty indicator of 1.5 and who are native English speakers. Do it using the proper transformed model.

11. Answer the previous question for female professors. Interpret the difference.

12. Verify the two intervals using the predict() function.

13. Answer the last three questions for the predicted evaluation of one professor with the same caracteristics.

Question 2:
For this question, load the data file CPS1985.rda that you will find in the Data folder of Learn. The following shows what we have in the dataset.

```
load("/Users/andres.arcila/Dropbox/Teaching/U Waterloo/Spring 2022/ECON322/R_Tutorial/Tutorial 6/CPS1985.rda")
data<-CPS1985
str(data)
```

```
## 'data.frame': 534 obs. of  11 variables:
##  $ wage      : num  5.1 4.95 6.67 4 7.5 ...
##  $ education : num  8 9 12 12 12 13 10 12 16 12 ...
##  $ experience: num  21 42 1 4 17 9 27 9 11 9 ...
##  $ age       : num  35 57 19 22 35 28 43 27 33 27 ...
##  $ ethnicity : Factor w/ 3 levels "cauc","hispanic",..: 2 1 1 1 1 1 1 1 1 1 ...
##  $ region    : Factor w/ 2 levels "south","other": 2 2 2 2 2 2 1 2 2 2 ...
##  $ gender    : Factor w/ 2 levels "male","female": 2 2 1 1 1 1 1 1 1 1 ...
##  $ occupation: Factor w/ 6 levels "worker","technical",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ sector    : Factor w/ 3 levels "manufacturing",..: 1 1 1 3 3 3 3 3 1 3 ...
##  $ union     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 2 1 1 1 1 ...
##  $ married   : Factor w/ 2 levels "no","yes": 2 2 1 1 2 1 1 1 2 1 ...
```

Estimate the following model for males and females (separately).

$$log(wage) \quad = \quad \beta_0 + \beta_1 education + \beta_2 experience + \beta_3 I(experience^2) + u_i$$

Then construct a 95% confidence interval for the number of years of experience that maximizes $\log(wage)$ for males and one for females. Use bootstrap to get the standard errors. To simplify the exercise create two subsamples, one for male and one for female:

```
datam <- subset(data, gender=="male")
dataf <- subset(data, gender=="female")
```