ECO 4421: Introduction to Econometric Methods
Department of Economics
Florida Atlantic University
Bonus Computer Project 2

You need do the project using R/Rstudio, compile the output into a (HTML/WORD/PDF) file. You may use "smart comments" to write your explanations. The output should contain both commands and results. Upload your output to Canvas.

The file `CPS12.csv` contains data for full-time workers ages 25-34, with a high school or bachelor diploma as their highest degree. In the following regressions, you will investigate the relationship between a worker's age and earnings. Generally, older workers have more job experience, leading to higher productivity and higher earnings.

(a) Use the `read.csv` command to read the `CPS12.csv` data set into R. Use the `attach` command to attach the data set into R.

(b) Print out an summary of the data set. In particular, find and report the sample average of the variables `ahe`, `age`, `female` and `bachelor`, respectively.

(c) Run a regression (`model1`) of average hourly earnings, `ahe` on `age`, `female` and `bachelor`. Find and use a sentence to interpret the meaning of the regression coefficients.

(d) Abby is a 28-years-old, female worker with a bachelor degree. Based on the regression from (`model1`), predict her average hourly earnings.

(e) Brian is a 32-years-old, male worker without a bachelor degree. Based on the regression from (`model1`), predict his average hourly earnings.

(f) Cindy is a 25-years-old, female worker without a bachelor degree. Based on the regression from (`model1`), predict her average hourly earnings.

(g) Find the SER, R^2 and the adjusted R^2 from the regression from (`model1`). Why are R^2 and the adjusted R^2 similar to each other in this regression?

(h) Run a regression (`model2`) of average hourly earnings, `ahe` on `age` only.

(i) Perform an F-test to compare (`model1`) and (`model2`). You may use the (`anova`) command.

(j) Based on the F-test result, do we prefer (`model1`) or (`model2`)?

(k) Run a regression (`model3`) of the logarithm of average hourly earnings, `log(ahe)` on `age`, age^2, `female` and `bachelor`.

(l) Run a regression (`model4`) of the logarithm of average hourly earnings, `log(ahe)` on `age`, age^2, `female`, $female{*}age$, $female{*}age^2$, and `bachelor`.

(m) Perform an F-test to compare (`model3`) and (`model4`). You may use the (`anova`) command.

(n) Based on the F-test result, do we prefer (`model3`) or (`model4`)?