Homework Assignments/Problem Set 3
Submission due: December 4th 2020 (Friday); 12:00NN.

Direction: Answer all the questions and submit your exercise in PDF. Make sure to include your R scripts or computer codes as an **appendix** to your submission (You can copy your codes from the R script and paste them on the MS Word document but **please do not include them as part of your answer**!). Please comment on each procedure so I would know what you are doing (or intend to do). Submission is via E-BART(clickj Quick links on the ELE page).

(20 points)

1. Females, on average, are shorter and weigh less than males. One of your friends, who is a pre-med student, tells you that in addition, females will weigh less for a given height. To test this hypothesis, you collect height and weight of 29 female and 81 male students at your university. A regression of the weight on a constant, height, and a binary variable, which takes a value of one for females and is zero otherwise, yields the following result:

$\widehat{Studentw}$ = -229.21 – 6.36 × *Female* + 5.58 × *Height*, R^2=0.50, *SER* = 20.99

where *Studentw* is weight measured in pounds and *Height* is measured in inches.

(a) Interpret the results. Does it make sense to have a negative intercept? (10 points)
(b) You decide that in order to give an interpretation to the intercept you should rescale the height variable. One possibility is to subtract 5 ft. or 60 inches from your *Height*, because the minimum height in your data set is 62 inches. The resulting new intercept is now 105.58. Can you interpret this number now? Do you think that the regression R^2 has changed? What about the standard error of the regression? (10 points)

(20 points)
2. You have collected data for 104 countries to address the difficult questions of the determinants for differences in the standard of living among the countries of the world. You recall from your macroeconomics lectures that the neoclassical growth model suggests that output per worker (per capita income) levels are determined by, among others, the saving rate and population growth rate. To test the predictions of this growth model, you run the following regression:

$\widehat{RelPersInc}$ = 0.339 – 12.894 × n + 1.397 × S_K, R^2=0.621, *SER* = 0.177
(0.068) (3.177) (0.229)

where *RelPersInc* is GDP per worker relative to the United States, n is the average population growth rate, 1980-1990, and S_K is the average investment share of GDP from 1960 to 1990 (remember investment equals saving). Numbers in parentheses are for heteroskedasticity-robust standard errors.
(a) Calculate the *t*-statistics and test whether or not each of the population parameters are significantly different from zero. (10 points)
(b) You remember that human capital in addition to physical capital also plays a role in determining the standard of living of a country. You therefore collect additional data on the average educational attainment in years for 1985, and add this variable (*Educ*) to the above regression. This results in the modified regression output:

$\widehat{RelPersInc} = 0.046 - 5.869 \times n + 0.738 \times S_K + 0.055 \times Educ$, R^2=0.775, $SER = 0.1377$

$(0.079)\quad(2.238)\quad(0.294)\quad(0.010)$

How has the inclusion of *Educ* affected your previous results? (10 points)

(60 points)

3. Lead is toxic, particularly for young children, and for this reason, government regulations severely restrict the amount of lead in our environment. But this was not always the case. In the early part of the 20th century, the underground water pipes in many U.S. cities contained lead, and lead from these pipes leached into drinking water. In this exercise, you will investigate the effect of these lead water pipes on infant mortality. On the ELE (assessment page), you will find the data file "lead_mortality.csv", which contains data on infant mortality, type of water pipes (lead or nonlead), water acidity (ph), and several demographic variables for 172 U.S. cities in 1900. A detailed description is found in https://wps.pearsoned.co.uk/wps/media/objects/16103/16489878/data3eu/Lead_Mortality_Description.pdf

a. Compute the average infant mortality rate (infrate) for cities with lead pipes and for cities with nonlead pipes. Is there a statistically significant difference in the averages? (10 points)
b. The amount of lead leached from lead pipes depends on the chemistry of the water running through the pipes. The more acidic the water is (that is, the lower its ph), the more lead is leached. Run a regression of infrate on lead, ph, and the interaction term lead * ph.
   i. The regression includes four coefficients (the intercept and the three coefficients multiplying the regressors). Explain what each coefficient measures. (10 points)
   ii. Plot the estimated regression function relating infrate to ph for Lead = 0 and for Lead = 1. Describe the differences in the regression functions, and relate these differences to the coefficients you discussed in (i). (10 points)
   iii. Does Lead have a statistically significant effect on infant mortality? Explain. Does the effect of Lead on infant mortality depend on ph? Is this dependence statistically significant? (10 points)
   iv. What is the average value of ph in the sample? At this ph level, what is the estimated effect of Lead on infant mortality? What is the standard deviation of ph? Suppose the ph level is one standard deviation lower than the average level of ph in the sample: What is the estimated effect of Lead on infant mortality? What if ph is one standard deviation higher than the average value?
c. The analysis in (b) may suffer from omitted variable bias because it neglects factors that affect infant mortality and that might potentially be correlated with lead and ph. Investigate this concern, using the other variables in the data set. What happened to the previous estimated slope coefficients after including other variables? (10 points)