E471: Econometric Theory and Practice I

Problem Set: R Exercises

Instructor:	Stefan Weiergraeber	Due:	Nov-17-2017
Email:	sweiergr@iu.edu		(in class)

Question 1

Suppose an econometrician models the relationship between X_1 (an individual's distance to the next university) and Y (the individual's years of education) as:

$$Y = 13.96 - 0.07X_1 + \epsilon$$

(a) Plot the expected relationship and interpret it.

(b) Another econometrician thinks one should model the relationship differently for low and high income individuals:

$$Y_p = 13.68 - 0.05X_{p,1} + \epsilon_p$$

for low income individuals and

$$Y_r = 14.57 - 0.08X_{r,1} + \epsilon_r$$

for high income individuals. Plot the two functions and interpret them.

(c) Now, let X_2 be an indicator for high income individuals. How can we summarize the equations for Y_p and Y_r in one equation?

(d) Open the following data set:
http://wps.aw.com/wps/media/objects/3254/3332253/datasets2e/datasets/CollegeDistance.dta
using the R-function `read.dta` contained in the R package `foreign`.[1] You can find a description of the data set at:
http://wps.aw.com/wps/media/objects/3254/3332253/datasets2e/datasets/CollegeDistance_DataDescription.pdf

(e) Find variables that correspond to Y, X_1 und X_2, and complement the existing graphs with a scatter plot of these variables.

(f) Many econometricians will criticize that the coefficients above were picked randomly. Estimate the model:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

using the methods of ordinary least squares based on the full sample. Estimate the model twice: once using R's `lm` command and a second time, programming the OLS estimator "by hand" using the formula $\hat{\beta} = (X^T X)^{-1} X^T Y$.

[1] Alternatively, you can use the `haven`-package.

(g) Estimate the model:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

separately for low and high income individuals using the methods of ordinary least squares. Estimate the model twice: once using R's `lm` command and a second time, programming the OLS estimator "by hand" using the formula $\hat{\beta} = (X^T X)^{-1} X^T Y$.

(h) Now estimate the following model:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1 X_2 + \eta$$

(i) Compare the estimates from (f), (g) and (h).

(j) Use R's `predict`-command and the model from part (h) to predict the expected outcomes separately for low and high income individuals and plot the relationship.