HED 612

Homework #11

**Directions**:

* Write your name on this document
* If the questions below ask you to execute R commands, then copy all R syntax (indicated via Courier Font) into the R script
* The R script should have a #comment indicating what number question the R syntax refers to for this assignment
* *Submit your answers file along with your R script to the D2L Dropbox*

**Before you begin**:

*Download the R Script for this homework assignment*

* I have already created this R script!
* Download it from on D2L: Under Lecture 11, Homework Assignment #11 subfolder
* On lines 28-80, I have cleaned and created analysis variables for you.
* You will need to run everything in this R script before you can answer the questions below!

*Understand how I created the variables in the R Script for this homework assignment*

* **The dependent variable is num\_coll\_apps**
  + I created this variable from the original variable called f2napp1p
  + All I did was recoded -9/-8/-4/-3 values in f2napp1p to NA in num\_coll\_apps
* **The independent variable of interest is first\_gen**
  + I define “first generation college student” as a student whose mother’s highest level of education is “did not finish high school” or “graduated from high school or GED”
  + I coded these two categories as the value 1 on first\_gen from the original variable f1mothed
  + All other categories from the original variable f1mothed were coded as 0 on first\_gen
  + Based on my coding for first\_gen:
    - Reference category = Non-first generation college students (first\_gen = 0)
    - Non- reference category = Firsts generation college studentss (first\_gen = 1)
* **There are two control variables: student\_motivation and ses\_quartile**
  + Student motivation is constructed from the original variable f1stexp which measures “how far in school a student thinks they will get.”
    - All I did was recoded -8/-4/values in f1stexp to NA in student\_motivation
    - So our student\_motivation values and labels are as follows:
      * 1 = Less than high school graduation
      * 2 = GED or other equivalency only
      * 3 = High school graduation only
      * 4 = Attend or complete 2-year college/school
      * 5 = Attend college, 4-year degree incomplete
      * 6= Graduate from college
      * 7 = Obtain Master's degree or equivalent
      * 8 = Obtain PhD, MD, or other advanced degree
      * 9 = Don't know
  + SES quartile is constructed from the original variable f1ses1qu which measures SES based on quartiles
    - All I did was recoded -8/-4/values in f1ses1qu to NA in ses\_quartile
    - So our ses\_quartile values and labels are as follows:
      * 1= lowest SES quartile
      * 2 = second SES quartile
      * 3= third SES quartile
      * 4= highest SES quartile

***For the next set of questions, we will investigate the effect of being a “first generation college student” on the number of colleges students apply to.***

1. Write out the population regression model forthe multivariate regression on the effect of being a “first generation college student” (X) on the number of colleges students apply to (Y) with ses\_quartile as our only control variable.

* Where:
  + If you remember from lectures 7 & 8, R creates the dummy variable shortcut for us when running the model, but the population regression model should be written out using all categories in each categorical variable as dummies
  + *Assume the lowest quartile is the reference group. In other words, this group should be left out of the regression*

1. Run the simple regression (only the independent variable of interest) in R for the effect of being a “first generation college student” (X) on the number of colleges students apply to (Y). *(Don’t forget to run the model with robust standard errors!)*
2. Run the multivariate regression (with ses\_quartile as the only control variable) in R for the effect of being a “first generation college student” (X) on the number of colleges students apply to (Y). *(Don’t forget to run the model with robust standard errors!)*
3. For the first generation college student coefficient in your *multivariate regression model* in Q3:
   * + - 1. Interpret the coefficient in words
         2. State whether the coefficient p-value is insignificant or significant at the .1, .05, .01 level.
4. Run the multivariate regression (with student\_motivation and ses\_quartile as control variables) in R for the effect of being a “first generation college student” (X) on the number of colleges students apply to (Y). *(Don’t forget to run the model with robust standard errors!)*
5. For the first generation college student coefficient in your *multivariate regression model* in Q5:
   1. Interpret the coefficient in words
   2. State whether the coefficient p-value is insignificant or significant at the .1, .05, .01 level.

***If you really feel like testing yourself, write out the population regression model for the multivariate regression in Q5***

* (Hint: Apart from , you should have 12 betas and 12 Xs because both our control variables are categorical variables)