HED 612

Homework #14

**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**:

*Create a new R Script for this homework assignment*

* Open the RProject you created last week (should be in your main hed612 folder)
* Once the RStudio window opens, within the R project session, open a new R Script
  + files >>> New File… >>> R Script
* Save the file as HW14\_lastname.R within lecture14 subfolder

*About the data*

The CA Schools dataset contains data on test performance, school characteristics and student demographic backgrounds. The data used here are from all 420 K-6 and K-8 districts in California with data available for 1998 and 1999. Each observation indicates one school district.

***Part 1: Continuous by Categorical Interaction***

*For the next set of questions, we will use* ***MA School Data*** *to investigate the effect of district average income per capita on test scores.*

* *Y=* testscr, *district average test score*
* *X =* str*, student teacher ratio*
* *Z =* ELL*, 0/1 District is High ELL district*
  + *0 = District is a “low” ELL district (less than 30% ELL students)*
  + *1= District is a “high” ELL district (greater than 30% ELL students)*

1. Within your R script: load the haven, tidyverse, labelled, and ggplot2 libraries; check that your directory is set to the R project; open the California data.

caschool<- read\_dta("data/ca/caschool-v2.dta")

1. Create a variable named ell that is a binary variable indicating districts have more than 30%.

caschool <- caschool %>% mutate(

ell = ifelse(el\_pct>=30, 1, 0)

)

1. Write out the population regression model for investigating whether the effect of student teacher ratio (X) on test scores differs (Y) by ELL district status (Z). (hint: run a regression model with a continuous (str) by categorical (ell) interaction).
2. Run the interaction regression model in R via the R shortcut (see R lecture R.script for syntax!).
   1. Write out the OLS prediction line without estimates
   2. Write out the OLS prediction line with estimates
3. Is there an interaction effect between student teacher ratio (X) and ELL status (Z)?
   1. State the null and alternative hypothesis being tested by
   2. Is the coefficient significant at an alpha level of 0.05?
   3. Can we reject the null hypothesis? (based on this answer the overall question #3).

*Whether or not the interaction coefficient is statistically significant in Q3, complete the remaining questions as if it is….*

1. What is the predicted value of test scores (Y) for “high ELL” districts (Z) with a student teacher ratio of 15 (X)?
2. What is the predicted value of test scores (Y) for “low ELL” districts (Z) with a student teacher ratio of 15 (X)?
3. What is the effect of X on Y for different values of Z…
   1. Interpret
   2. Interpret
   3. Interpret
   4. How do we calculate the average effect of student teacher ratio (X) on test scores (Y) for “high ELL” districts (Z=1)? Calculate and interpret this effect.
   5. (You don’t need to answer the overall question #6).