HED 612

Homework #12

**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 CA Data [If you did not complete this during class]*:

* Create a new data folder called “ca”
  + hed612 >>> data >>> ca
* Download the California Dataset from D2L (under Datasets)
  + Place the “caschool-v2” dataset into the “ca” folder you created in the previous step

*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 HW12\_lastname.R within lecture12 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.

***For the next set of questions, we will investigate the effect of district average percent of students qualifying for CalWORKs on district average student math scores.***

* *CalWORKs is a public assistance program that provides cash aid and services to eligible families that have a child(ren) in the home*

1. Explain, in your own words or using an example, what the nonlinear function “*the effect of on Y depends on .”*
2. Investigate the effect of district average percent of students qualifying for CalWORKs (X= calw\_pct) on district average student math scores (Y= math\_scr)by creating a scatterplot of the data. Add a linear model to the scatterplot. By just looking at the scatterplot, does the linear model fit the data well? Why or why not? (hint: use R syntax from the Lecture 12 R script to create the plot)
3. Write out the population regression model for the effect of district average percent of students qualifying for CalWORKs (X= calw\_pct) on district average student math scores (Y= math\_scr)**using a quadratic of X**.
4. Run the regression for the population regression model in #3. (hint: you need to use the I() function and ^ operator within your lm() function).
   1. Write out the OLS prediction line with estimates.
5. For the regression in Q4:
   * + - 1. What does the coefficient represent now that our model includes a quadratic term of X?
         2. Calculate the rate of change in , district average math test scores, for a change in **9 to 10 percent** in district average percent of students qualifying for CalWORKs. (Calculate by hand and use predict() to check your work)
         3. Calculate the rate of change in , district average math test scores, for a change in **30 to 31 percent** in district average percent of students qualifying for CalWORKs. (Calculate by hand and use predict() to check your work)
6. For the regression in Q4:
   1. What does the coefficient test in regards to the quadratic term of X? (hint: state the null and alternative hypotheses)
   2. State whether the coefficient p-value is insignificant or significant at the .1, .05, .01 level.
   3. What does the direction (positive or negative) of the coefficient tell us about our data?
   4. What does the magnitude of the coefficient tell us about our data?