HED 612

Homework #6

**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 HW4\_lastname.R within lecture2 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.

# Questions related to caschools-v2.dta

Our research question for this homework assignment is: What is the effect of number of computers (X) on district average test scores (Y)?

1. Within the R script created above: 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. Write down the population linear regression model for the research question above and label X, Y, population intercept, population regression coefficient, and error term.

Where:

* = district average student test scores
* = number of computers
* = “population intercept”: average Y for observation with X=0
* = “population regression coefficient”: the average effect of a one-unit increase in X on the value of Y
* = “residual”: all other variables not included in the model that affect the value of Y

1. Write down the null and alternative hypotheses about the effect of number of computers on district average test scores.
2. In R, run a regression of the effect of number of computers (computer) on test scores (testscr)

mod1 <- lm(testscr ~ computer, data=caschool)

summary(mod1)

1. Write out the formula for t-statistic of . Using the regression output from the previous question, show that t = -1.512.
2. Draw the sampling distribution of under the assumption that the null hypothesis is true.
   1. What does each observation represent? Each observation represents a one random sample beta regression coefficient
   2. Label the value of the population regression coefficient associated with the null hypothesis.
   3. Label the sample estimate of the population regression coefficient, and the observed t-value
   4. Shade the p-value regions associated with this t-value

