1. Download the hw\_w2\_1.RDat file from canvas and load it into R

load(file = 'hw\_w2\_1.RDat')

## look at dataframe

head(hw\_w2\_1)

Run a linear regression and check whether there is suppression, and if so, what type?

2. Download the hw\_w2\_2.RDat file from canvas and load it into R

load(file = 'hw\_w2\_2.RDat')

Fit a quadratic model and center at the beginning of the study, at the grand-mean and at the end.

Plot the figure with the fitted line and the simple slope or instantaneous rate of change at the three centering points.

3. In this exercise you will simulate two variables that are statistically independent of each other to see what happens when we run a regression of one on the other.

1. First generate 1000 data points from a normal distribution with mean 0 and standard deviation 1 by typing var1 <- rnorm(1000,0,1) in R. Generate another variable in the same way (call it var2). Run a regression of one variable on the other. Is the slope coefficient statistically significant?
2. Now run a simulation repeating this process 100 times. This can be done using a loop. From each simulation, save the z-score (the estimated coefficient of var1 divided by its standard error). If the absolute value of the z-score exceeds 2, the estimate is statistically significant. Here is code to perform the simulation:  
     
   z.scores <- rep (NA, 100)

for (k in 1:100) {

var1 <- rnorm (1000,0,1)

var2 <- rnorm (1000,0,1)

fit <- lm (var2 ~ var1)

z.scores[k] <- coef(fit)[2]/se.coef(fit)[2]

}

How many of these 100 z-scores are statistically significant?

4. Download the child.iq.dta dataset. You have access to children’s test scores at age 3, mother’seducation, and the mother’s age at the time she gave birth for a sample of 400

children. The data are a Stata file which you can read into R by saving in your working directory and then typing the following:

library ("foreign")

iq.data <- read.dta ("child.iq.dta")

1. Fit a regression of child test scores on mother’s age, display the data and fitted model, check assumptions, and interpret the slope coefficient. When do you recommend mothers should give birth? What are you assuming in making these recommendations?
2. Repeat this for a regression that further includes mother’s education, interpreting both slope coefficients in this model. Have your conclusions about the timing of birth changed?

5. Download data from Hamermesh and Parker (2005) on student

evaluations of instructors’ beauty and teaching quality for several courses at the

University of Texas. The teaching evaluations were conducted at the end of the

semester, and the beauty judgments were made later, by six students who had

not attended the classes and were not aware of the course evaluations.

load(file = './TeachingRatings.rda')

1. Run a regression using beauty to predict course evaluations (eval), controlling for various other inputs. Display the fitted model graphically, and explaining the meaning of each of the coefficients. Plot the residuals versus fitted values.
2. Fit some other models, including beauty and also other input variables. Consider at least one model with interactions. For each model, state what the predictors are, and what the inputs are (see Section 2.1), and explain the meaning of each of its coefficients.

6 The data wfw90.dat is from the Work, Family, and Well-Being Survey (Ross, 1990). Pull out the data on earnings, sex, height, and weight.

1. Fit a linear regression model predicting earnings from height. What transformation should you perform in order to interpret the intercept from this model as average earnings for people with average height?
2. Fit some regression models with the goal of predicting earnings from some combination of sex, height, and weight. Be sure to try various transformations and interactions that might make sense. Choose your preferred model and justify.
3. Interpret all model coefficients.

7. An economist runs a regression examining the relations between the average price

of cigarettes, P , and the quantity purchased, Q, across a large sample of counties

in the United States, assuming the following functional form, log Q = α+β log P .

Suppose the estimate for β is 0.3. Interpret this coefficient.