**BACS - HW 12**

Let’s take another look at interactions in our cars dataset. For this week, let’s only use the following data:

1. **mpg: miles-per-gallon (dependent variable)**
2. weight: weight of car
3. acceleration: acceleration ability of car
4. model\_year: year model was released
5. origin: place car was designed (1: USA, 2: Europe, 3: Japan)
6. cylinders: cylinders in engine *(only used in Question 3)*

Create a data.frame called cars\_log with log-transformed columns for mpg, weight, and acceleration  
(model\_year and origin don’t have to be transformed)

**Question 1)** Let’s visualize how weight and acceleration are related to mpg.

1. Let’s visualize how weight might *moderate* the relationship between acceleration and mpg:
   1. Create two *subsets* of your data, one for light-weight cars (less than mean weight)   
      and one for heavy cars (higher than the mean weight)  
      *HINT: consider carefully how you compare log weights to mean weight*
   2. Create a *single* scatter plot of acceleration vs. mpg, with different colors and/or shapes for light versus heavy cars
   3. Draw two slopes of acceleration-vs-mpg over the scatter plot:   
      one slope for light cars and one slope for heavy cars (distinguish them by appearance)
2. Report the full summaries of two separate regressions for light and heavy cars where  
    log.mpg. is dependent on log.weight., log.acceleration., model\_year and origin
3. *(not graded)* Using your intuition only: What do you observe about light versus heavy cars so far?

**Question 2)** Using the fully transformed dataset from above (cars\_log), to test whether we have moderation.

1. *(not graded)* Between weight and acceleration ability (in seconds), use your intuition and experience to state which variable might be a moderating versus independent variable, in affecting mileage.
2. Use various regression models to model the possible moderation on log.mpg.:  
   (use log.weight., log.acceleration., model\_year and origin as independent variables)
   1. Report a regression *without any interaction terms*
   2. Report a regression *with an interaction between weight and acceleration*
   3. Report a regression *with a mean-centered interaction term*
   4. Report a regression *with an orthogonalized interaction term*
3. For each of the interaction term strategies above (raw, mean-centered, orthogonalized) what is the correlation between that interaction term and the two variables that you multiplied together?

*(see Question 3 on next page)*

****Question 3)** We saw earlier that the number of cylinders does not seem to *directly* influence mpg when car weight is also considered. But might cylinders have an *indirect* relationship with mpg through its weight?   
  
Let’s check whether weight *mediates* the relationship between cylinders and mpg, even when other factors are controlled for. Use log.mpg., log.weight., and log.cylinders as your main variables, and keep log.acceleration., model\_year, and origin as *control variables* (see gray variables in diagram).

1. Let’s try computing the direct effects first:
   1. Model 1: Regress **log.weight.** over **log.cylinders.** only  
      *(check whether number of cylinders has a significant direct effect on weight)*
   2. Model 2: Regress **log.mpg.** over **log.weight.** and all control variables  
      *(check whether weight has a significant direct effect on mpg with other variables statistically controlled?)*
2. What is the *indirect effect* of cylinders on mpg? *(use the product of slopes between model 1 & 2)*
3. Let’s bootstrap for the confidence interval of the *indirect effect* of cylinders on mpg
   1. Bootstrap regression models 1 & 2, and compute the indirect effect each time:   
      what is its 95% CI of the *indirect effect* of log.cylinders. on log.mpg.?
   2. Show a density plot of the distribution of the 95% CI of the indirect effect