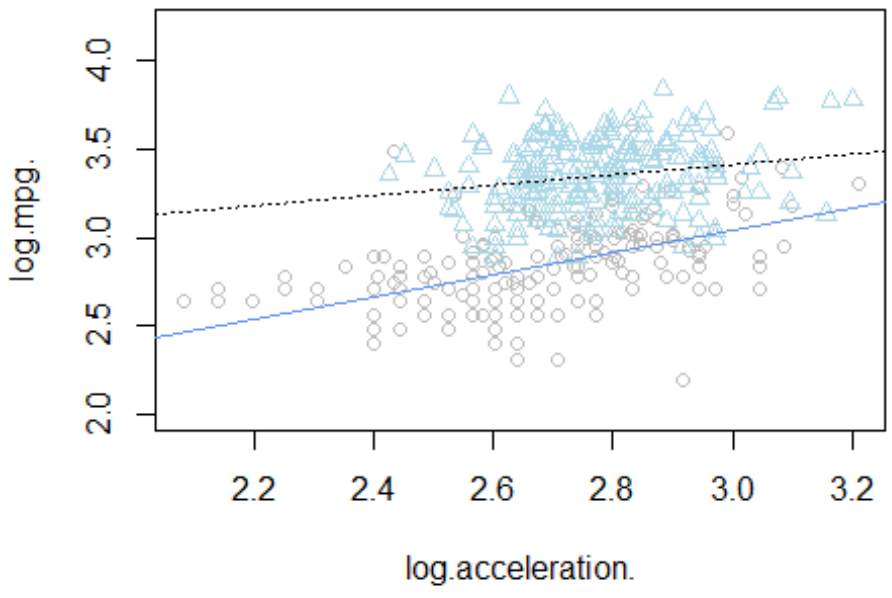
HW13_109071503

Question 1: nonlinearity and log-transforms

a. Visualize how weight might moderate the relationship between acceleration and mpg

i. Create two subsets of your data
ii. Create a single scatter plot of acceleration vs. mpg.
iii. Draw two slopes of acceleration-vs-mpg over the scatter plot

```
# (i)
cars <- read.table("auto-data.txt", header=FALSE, na.strings = "?")
names(cars) <- c("mpg", "cylinders", "displacement", "horsepower", "weight",
"acceleration", "model_year", "origin", "car_name")
cars_log <- with(cars, data.frame(log(mpg), log(weight), log(acceleration), model_year, o
rigin))
light_cars = subset(cars_log, log.weight. < mean(cars_log[, "log.weight."]))
heavy_cars = subset(cars_log, log.weight. > mean(cars_log[, "log.weight."]))
# (ii)
light_regr = lm(formula = log.mpg. ~ log.acceleration., data = light_cars)
heavy_regr = lm(formula = log.mpg. ~ log.acceleration., data = heavy_cars)
with(heavy_cars, plot(log.acceleration., log.mpg., col="gray", ylim = c(2, 4.2)))
with(light_cars, points(log.acceleration., log.mpg., pch=2, col="lightblue"))
# (iii)
abline(light_regr, lty="dotted")
abline(heavy_regr, lty="solid", col="cornflowerblue")
```



b. Report the full summaries of two separate regressions for light and heavy cars

```
lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year + factor(origin), da
ta = light_cars)
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin), data = light_cars)
##
## Coefficients:
##       (Intercept)        log.weight.  log.acceleration.         model_year
##           6.80901           -0.82195            0.11114            0.03334
##   factor(origin)2    factor(origin)3
##           0.04231            0.02092
```

```
lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year + factor(origin), da
ta = heavy_cars)
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin), data = heavy_cars)
##
## Coefficients:
##       (Intercept)        log.weight.  log.acceleration.         model_year
##           7.13289           -0.82552            0.03122            0.03173
##   factor(origin)2    factor(origin)3
##           0.09903            0.06315
```

c. (not graded) What do you observe about light versus heavy cars so far? The coefficient of log.acceleration. of two subset is different.

Question 2: Using the fully transformed dataset from above (cars_log), to test whether we have moderation.

a. (not graded) Between weight and acceleration ability, use your intuition and experience to state which variable might be a moderating versus independent variable, in affecting mileage.

b. Use various regression models to model the possible moderation on log.mpg

i. Report a regression without any interaction terms

```
full_regr <- with(cars_log, lm(log.mpg.~ log.weight.+ log.acceleration.+ model_year + fac
tor(origin)))
summary(full_regr)
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38275 -0.07032  0.00491  0.06470  0.39913
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.431155   0.312248  23.799  < 2e-16 ***
## log.weight.       -0.876608   0.028697 -30.547  < 2e-16 ***
## log.acceleration.  0.051508   0.036652   1.405  0.16072
## model_year         0.032734   0.001696  19.306  < 2e-16 ***
```

```
## factor(origin)2    0.057991   0.017885   3.242  0.00129 **
## factor(origin)3    0.032333   0.018279   1.769  0.07770 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1156 on 392 degrees of freedom
## Multiple R-squared:  0.8856, Adjusted R-squared:  0.8841
## F-statistic: 606.8 on 5 and 392 DF,  p-value: < 2.2e-16
```

ii. Report a regression with an interaction between weight and acceleration

```
regr_interaction <- with(cars_log,lm(log.mpg.~log.weight.+log.acceleration.+log.weight.*l
og.acceleration.+ model_year+ factor(origin)))
summary(regr_interaction)

##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + log.weight. *
##     log.acceleration. + model_year + factor(origin))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37807 -0.06868  0.00463  0.06891  0.39857
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.089642   2.752872   0.396  0.69245
## log.weight.                  -0.096632   0.337637  -0.286  0.77488
## log.acceleration.             2.357574   0.995349   2.369  0.01834 *
## model_year                    0.033685   0.001735  19.411  < 2e-16 ***
## factor(origin)2               0.058737   0.017789   3.302  0.00105 **
## factor(origin)3               0.028179   0.018266   1.543  0.12370
## log.weight.:log.acceleration. -0.287170   0.123866  -2.318  0.02094 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.115 on 391 degrees of freedom
## Multiple R-squared:  0.8871, Adjusted R-squared:  0.8854
## F-statistic: 512.2 on 6 and 391 DF,  p-value: < 2.2e-16
```

iii. Report a regression with a interaction term

```
log.weight._mc <- scale(cars_log$log.weight., center = TRUE, scale = FALSE)
log.acceleration._mc <- scale(cars_log$log.acceleration., center = TRUE, scale = FALSE)
#Mean-centered regression with interaction
regr_meancenter = lm(cars_log$log.mpg.~ log.weight._mc + log.acceleration._mc
+ log.weight._mc*log.acceleration._mc + cars_log$model_year
+ factor(cars_log$origin))
summary(regr_meancenter)

##
## Call:
## lm(formula = cars_log$log.mpg. ~ log.weight._mc + log.acceleration._mc +
##     log.weight._mc * log.acceleration._mc + cars_log$model_year +
##     factor(cars_log$origin))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.37807 -0.06868  0.00463  0.06891  0.39857
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.518882   0.132944   3.903 0.000112 ***
## log.weight._mc                    -0.880393   0.028585 -30.799  < 2e-16 ***
## log.acceleration._mc               0.072596   0.037567   1.932 0.054031 .
## cars_log$model_year                0.033685   0.001735  19.411  < 2e-16 ***
## factor(cars_log$origin)2           0.058737   0.017789   3.302 0.001049 **
## factor(cars_log$origin)3           0.028179   0.018266   1.543 0.123704
## log.weight._mc:log.acceleration._mc -0.287170   0.123866  -2.318 0.020943 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.115 on 391 degrees of freedom
## Multiple R-squared:  0.8871, Adjusted R-squared:  0.8854
## F-statistic: 512.2 on 6 and 391 DF,  p-value: < 2.2e-16
```

iv. Report a regression with an orthogonalized interaction term

```
acc_x_wt <- cars_log$log.acceleration. * cars_log$log.weight.
interaction_regr <- lm(acc_x_wt ~ cars_log$log.acceleration. + cars_log$log.weight.)
interaction_ortho <- interaction_regr$residuals
#Regression model with residual
regr_ortho = with(cars_log,lm(log.mpg.~ log.weight. + log.acceleration. + interaction_ort
ho
+ model_year + factor(origin)))
summary(regr_ortho)

##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + interaction_ortho +
##     model_year + factor(origin))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37807 -0.06868  0.00463  0.06891  0.39857
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.377176   0.311392  23.691  < 2e-16 ***
## log.weight.       -0.876967   0.028539 -30.729  < 2e-16 ***
## log.acceleration.  0.046100   0.036524   1.262  0.20764
## interaction_ortho -0.287170   0.123866  -2.318  0.02094 *
## model_year         0.033685   0.001735  19.411  < 2e-16 ***
## factor(origin)2    0.058737   0.017789   3.302  0.00105 **
## factor(origin)3    0.028179   0.018266   1.543  0.12370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.115 on 391 degrees of freedom
## Multiple R-squared:  0.8871, Adjusted R-squared:  0.8854
## F-statistic: 512.2 on 6 and 391 DF,  p-value: < 2.2e-16
```

c. 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?

```r
cars_log2 <- with(cars, data.frame(log(weight), log(acceleration)))
round(cor(cbind(cars_log2, acc_x_wt)),2)
```

```
##                   log.weight. log.acceleration. acc_x_wt
## log.weight.              1.00             -0.43     0.11
## log.acceleration.       -0.43              1.00     0.85
## acc_x_wt                 0.11              0.85     1.00
```

```r
round(cor(cbind(log.weight._mc, log.acceleration._mc, log.weight._mc*log.acceleration._m
c)),2)
```

```
##       [,1]  [,2]  [,3]
## [1,]  1.00 -0.43 -0.20
## [2,] -0.43  1.00  0.35
## [3,] -0.20  0.35  1.00
```

```r
round(cor(cbind(cars_log2, interaction_ortho)),2)
```

```
##                   log.weight. log.acceleration. interaction_ortho
## log.weight.              1.00             -0.43                 0
## log.acceleration.       -0.43              1.00                 0
## interaction_ortho        0.00              0.00                 1
```