BACS HW (Week13)

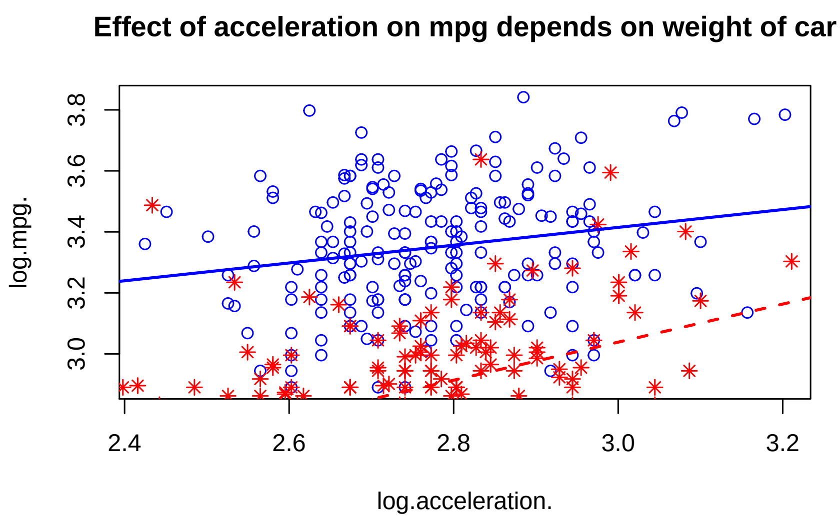
106070038

2021-05-23

Question 1 Let's visualize how weight and acceleration are related to mpg.

a. Let's visualize how weight might moderate the relationship between acceleration and mpg

```
# install.packages("logr")
library(logr)

cars <- read.table("auto-data.txt",)
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, origin))

# calculate the mean of log.weight.
mean_log_weight <- mean(cars_log$log.weight.)

# subset cars dataset by mean weight
cars_light <- subset(cars_log, log.weight. < mean_log_weight)
cars_heavy <- subset(cars_log, log.weight. > mean_log_weight)

# separate regressions of acceleration vs. mpg by mean weight
acc_regr_light <- lm(log.mpg. ~ log.acceleration., data=cars_light)
acc_regr_heavy <- lm(log.mpg. ~ log.acceleration., data=cars_heavy)

# plot the points
with(cars_light, plot(log.acceleration., log.mpg., pch=1, col ="blue", lwd=1,
                      main = "Effect of acceleration on mpg depends on weight of car"))
with(cars_heavy, points(log.acceleration., log.mpg., pch=8, col ="red", lwd=1))

# plot separate regression lines colorized by origin
abline(acc_regr_light, col = "blue", lwd=2, lty=1)
abline(acc_regr_heavy, col = "red", lwd=2, lty=2)
```



b. 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

```
regr_light <- lm(log.mpg. ~ ., data = cars_light)
summary(regr_light)
```

```
##
## Call:
## lm(formula = log.mpg. ~ ., data = cars_light)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37798 -0.07041 -0.00001  0.06714  0.30909
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.019008   0.591532  11.866   <2e-16 ***
## log.weight.       -0.840576   0.065252 -12.882   <2e-16 ***
## log.acceleration.  0.107638   0.058568   1.838   0.0676 .
## model_year         0.032605   0.002016  16.169   <2e-16 ***
## origin             0.009573   0.009642   0.993   0.3220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1107 on 200 degrees of freedom
## Multiple R-squared:  0.7048, Adjusted R-squared:  0.6989
## F-statistic: 119.4 on 4 and 200 DF,  p-value: < 2.2e-16
```

```
regr_heavy <- lm(log.mpg. ~ ., data = cars_heavy)
summary(regr_heavy)
```

```
##
## Call:
## lm(formula = log.mpg. ~ ., data = cars_heavy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36723 -0.07194  0.00062  0.06660  0.42834
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.037532   0.684986  10.274  < 2e-16 ***
## log.weight.        -0.822236   0.068282 -12.042  < 2e-16 ***
## log.acceleration.  0.056971   0.052844   1.078  0.28237
## model_year          0.030895   0.003215   9.610  < 2e-16 ***
## origin              0.064136   0.024397   2.629  0.00928 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1216 on 188 degrees of freedom
## Multiple R-squared:  0.7556, Adjusted R-squared:  0.7504
## F-statistic: 145.3 on 4 and 188 DF,  p-value: < 2.2e-16
```

c. (not graded) Using your intuition only: What do you observe about light versus heavy cars so far?

- Answer:
  - heavy car(0.7556) has bigger R-squared than light car(0.7048)
  - the effect of origin in linear model in heavy car is more significant than in light car

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.

- Answer: I think is **weight**.

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

i. Report a regression without any interaction terms

```
summary(lm(log.mpg. ~ log.weight. + log.acceleration., data=cars_log))
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration., data = cars_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48030 -0.09642 -0.01185  0.09372  0.56878
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       10.48669    0.33484  31.319  < 2e-16 ***
## log.weight.       -1.00048    0.03192 -31.345  < 2e-16 ***
## log.acceleration.  0.21084    0.04957   4.253 2.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1615 on 395 degrees of freedom
## Multiple R-squared:  0.775,  Adjusted R-squared:  0.7739
## F-statistic: 680.3 on 2 and 395 DF,  p-value: < 2.2e-16
```

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

```
cor(cars_log$log.weight., cars_log$log.weight.*cars_log$log.acceleration.)
```

```
## [1] 0.1083055
```

```
summary(lm(log.mpg. ~ log.weight. + log.acceleration. + log.weight.*log.acceleration., data=cars_log))
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + log.weight. *
##     log.acceleration., data = cars_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49728 -0.10145 -0.01102  0.09665  0.56416
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    16.0249     3.6950   4.337 1.84e-05 ***
## log.weight.                    -1.6878     0.4578  -3.687 0.000259 ***
## log.acceleration.              -1.8252     1.3537  -1.348 0.178351
## log.weight.:log.acceleration.   0.2529     0.1681   1.505 0.133123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1613 on 394 degrees of freedom
## Multiple R-squared:  0.7763, Adjusted R-squared:  0.7746
## F-statistic: 455.7 on 3 and 394 DF,  p-value: < 2.2e-16
```

iii. Report a regression with a mean-centered interaction term

```
weight_mc <- scale(cars_log$log.weight., center=TRUE, scale=FALSE)
acceleration_mc <- scale(cars_log$log.acceleration., center=TRUE, scale=FALSE)

cor(weight_mc, weight_mc*acceleration_mc)
```

```
##            [,1]
## [1,] -0.2026948
```

```
summary(lm(log.mpg. ~ weight_mc + acceleration_mc + weight_mc*acceleration_mc, data = cars_log))
```

```
##
## Call:
## lm(formula = log.mpg. ~ weight_mc + acceleration_mc + weight_mc *
##     acceleration_mc, data = cars_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49728 -0.10145 -0.01102  0.09665  0.56416
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3.106831   0.008857 350.791  < 2e-16 ***
## weight_mc                 -0.997466   0.031930 -31.239  < 2e-16 ***
## acceleration_mc            0.187500   0.051862   3.615 0.000339 ***
## weight_mc:acceleration_mc  0.252948   0.168071   1.505 0.133123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1613 on 394 degrees of freedom
## Multiple R-squared:  0.7763, Adjusted R-squared:  0.7746
## F-statistic: 455.7 on 3 and 394 DF,  p-value: < 2.2e-16
```

iv. Report a regression with an orthogonalized interaction term

```
wei_x_acc <- cars_log$log.weight. * cars_log$log.acceleration.
interaction_regr <- lm(wei_x_acc ~ cars_log$log.weight. + cars_log$log.acceleration.)
interaction_ortho <- interaction_regr$residuals

round(cor(cbind(cars_log, interaction_ortho)), 2)
```

```
##                    log.mpg. log.weight. log.acceleration. model_year origin
## log.mpg.               1.00       -0.87              0.46       0.58   0.56
## log.weight.           -0.87        1.00             -0.43      -0.28  -0.60
## log.acceleration.      0.46       -0.43              1.00       0.31   0.22
## model_year             0.58       -0.28              0.31       1.00   0.18
## origin                 0.56       -0.60              0.22       0.18   1.00
## interaction_ortho      0.04        0.00              0.00       0.21  -0.07
##                    interaction_ortho
## log.mpg.                        0.04
## log.weight.                     0.00
## log.acceleration.               0.00
## model_year                      0.21
## origin                         -0.07
## interaction_ortho               1.00
```

```r
summary(lm(log.mpg. ~ cars_log$log.weight. + cars_log$log.acceleration.
           + interaction_ortho, data = cars_log))
```

```
##
## Call:
## lm(formula = log.mpg. ~ cars_log$log.weight. + cars_log$log.acceleration. +
##     interaction_ortho, data = cars_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49728 -0.10145 -0.01102  0.09665  0.56416
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 10.48669    0.33430  31.369  < 2e-16 ***
## cars_log$log.weight.        -1.00048    0.03187 -31.395  < 2e-16 ***
## cars_log$log.acceleration.   0.21084    0.04949   4.260 2.56e-05 ***
## interaction_ortho            0.25295    0.16807   1.505    0.133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1613 on 394 degrees of freedom
## Multiple R-squared:  0.7763, Adjusted R-squared:  0.7746
## F-statistic: 455.7 on 3 and 394 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?

Raw

```
cor(cars_log$log.weight., cars_log$log.weight.*cars_log$log.acceleration.)

## [1] 0.1083055

cor(cars_log$log.acceleration., cars_log$log.weight.*cars_log$log.acceleration.)

## [1] 0.852881
```

- Answer: **log.weight.** is the moderator. The effect of log.acceleration. is contingent on log.weight..

Mean-centered

```
cor(acceleration_mc, weight_mc*acceleration_mc)

##           [,1]
## [1,] 0.3512271

cor(weight_mc, weight_mc*acceleration_mc)

##            [,1]
## [1,] -0.2026948
```

- Answer: Correlation isn't effected by centering. As a result, the correlation is **same as raw**.

Orthogonalized

```
round(cor(cbind(cars_log, interaction_ortho)), 2)

##                   log.mpg. log.weight. log.acceleration. model_year origin
## log.mpg.              1.00       -0.87              0.46       0.58   0.56
## log.weight.          -0.87        1.00             -0.43      -0.28  -0.60
## log.acceleration.     0.46       -0.43              1.00       0.31   0.22
## model_year            0.58       -0.28              0.31       1.00   0.18
## origin                0.56       -0.60              0.22       0.18   1.00
## interaction_ortho     0.04        0.00              0.00       0.21  -0.07
##                   interaction_ortho
## log.mpg.                       0.04
## log.weight.                    0.00
## log.acceleration.              0.00
## model_year                     0.21
## origin                        -0.07
## interaction_ortho              1.00
```

- Answer: The correlation between interaction_ortho and log.weight. is 0. And, the correlation between interaction_ortho and log.acceleration. is also 0.