HW13

106022113

5/18/2021

Question 1. Visualize how wieght and acceleration are related to mpg

a. Visualize how wieght moderate the relationship between acceleration and mpg

```
auto <- read.table("auto-data.txt",header = FALSE, na.strings = "?", stringsAsFactors = F)
names(auto) <- c("mpg","cylinders","displacement","horsepower","weight","acceleration","model_year","ori
cars_log <- with(auto, data.frame(log(mpg),log(cylinders),log(displacement),log(horsepower),log(weight)
```

```
light_car <- subset(cars_log, log.weight. < log(mean(auto$weight)))
heavy_car <- subset(cars_log, log.weight. > log(mean(auto$weight)))
```

i. Create two subsetss, one for light cars and one for heavy cars

```
library(dplyr)
```

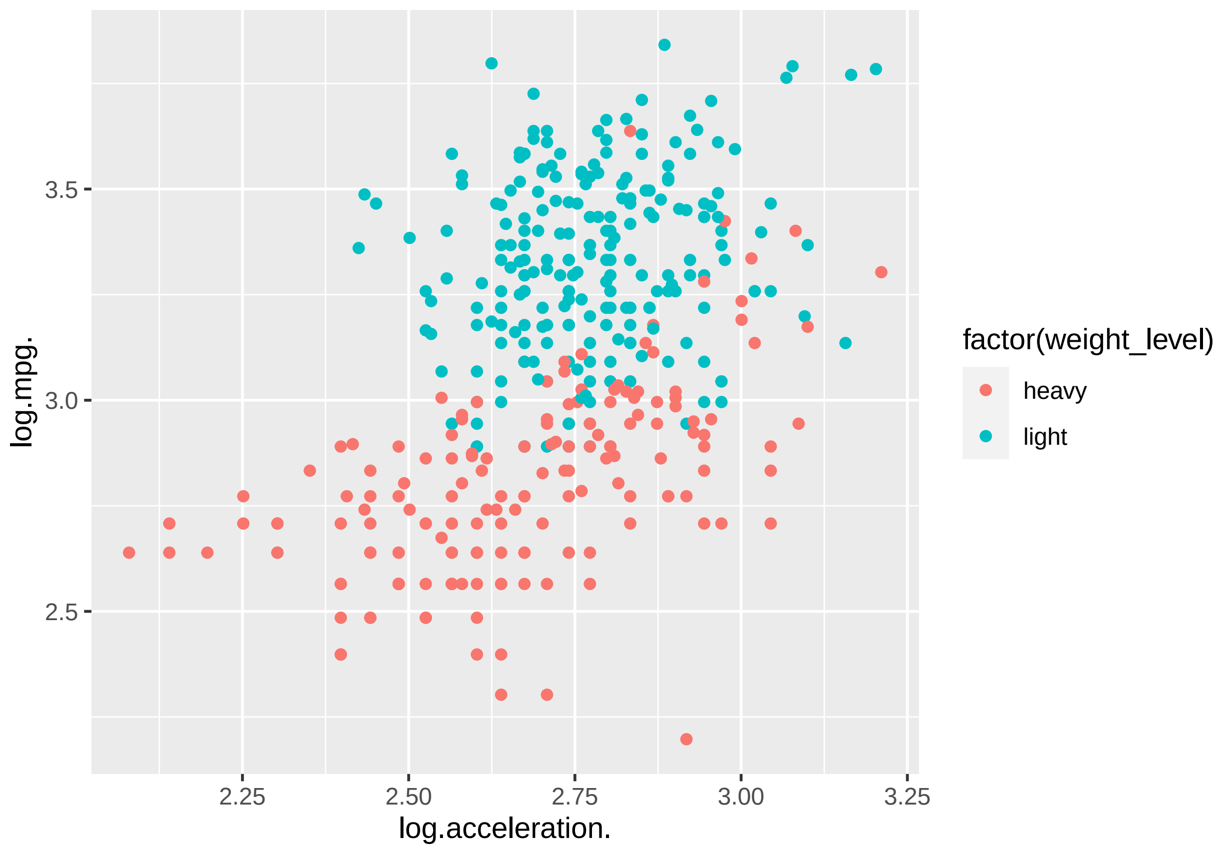
ii. Create scatter plot of acceleration vs mpg

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

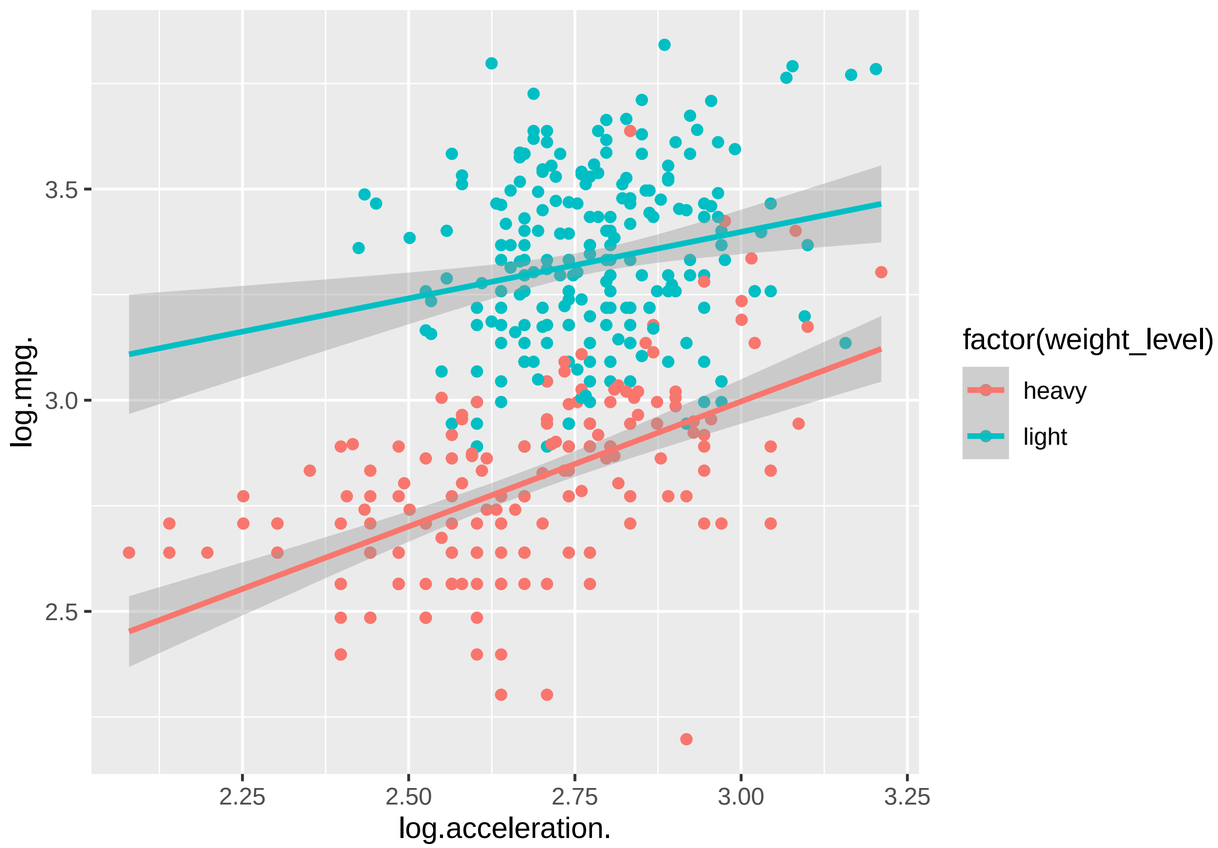
```
library(ggplot2)
cars_log <- cars_log %>% mutate(weight_level = ifelse(log.weight.>= log(mean(auto$weight)),"heavy", "lig
sctplt <- ggplot(data = cars_log, aes(x = log.acceleration., y=log.mpg., col = factor(weight_level)))+
  geom_point()
sctplt
```




```
sctplt+geom_smooth(method = lm,fullrange = TRUE)
```

iii. Draw two slopes of acceleration vs mpg over scatter plot

```
## 'geom_smooth()' using formula 'y ~ x'
```



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

```
light_regr <- lm(data = light_car, log.mpg.~ log.weight. + log.acceleration. + model_year + factor(orig
heavy_regr <- lm(data = heavy_car, log.mpg.~ log.weight. + log.acceleration. + model_year + factor(orig
summary(light_regr)
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin), data = light_car)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36464 -0.07181  0.00349  0.06273  0.31339
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.86661    0.52767  13.013   <2e-16 ***
## log.weight.        -0.83437    0.05662 -14.737   <2e-16 ***
## log.acceleration.   0.10956    0.05630   1.946   0.0529 .
## model_year          0.03383    0.00198  17.079   <2e-16 ***
## factor(origin)2     0.05129    0.01980   2.590   0.0102 *
## factor(origin)3     0.02621    0.01846   1.420   0.1571
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1112 on 221 degrees of freedom
## Multiple R-squared:  0.7292, Adjusted R-squared:  0.7231
## F-statistic:   119 on 5 and 221 DF,  p-value: < 2.2e-16
```

```
summary(heavy_regr)
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin), data = heavy_car)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36811 -0.06937  0.00607  0.06969  0.43736
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.188679   0.759983   9.459  < 2e-16 ***
## log.weight.       -0.822352   0.077206 -10.651  < 2e-16 ***
## log.acceleration.  0.040140   0.057380   0.700   0.4852
## model_year         0.030317   0.003573   8.486 1.14e-14 ***
## factor(origin)2    0.091641   0.040392   2.269   0.0246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1212 on 166 degrees of freedom
## Multiple R-squared:  0.7179, Adjusted R-squared:  0.7111
## F-statistic: 105.6 on 4 and 166 DF,  p-value: < 2.2e-16
```

c. Use intuition: what do you observe about the light v.s. heavy cars?

ANSWER: Normally, lighter cars will have better fuel efficiency performance than heavy cars. And we can observe that with mpg vs acceleration, lighter cars occupy mostly of the upper area. Also, the regression mpg intercept of light cars at the y axis is higher than heavy cars.

Question 2. Using the cars_log data to test moderation

a. Use tuition to state which might be a moderating versus independent variable

ANSWER: Acceleration might be a moderating variable.

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

```
full_regr <- lm(log.mpg.~ log.weight.+log.acceleration.+model_year+factor(origin),data = cars_log)
summary(full_regr)
```

i.Report regression without any interaction

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin), data = cars_log)
##
## 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
```

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

ii. Report regression with interaction between weight and acceleration

```
##
## 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
```

```
log.weight.mc <- scale(cars_log$log.weight.,center = TRUE, scale = FALSE)
log.acc.mc <- scale(cars_log$log.acceleration.,center = TRUE,scale = FALSE)
log.mpg.mc <- scale(cars_log$log.mpg., center = TRUE, scale = FALSE)
mc_regr <- lm(log.mpg.mc~ log.acc.mc+log.weight.mc+log.acc.mc*log.weight.mc)
summary(mc_regr)
```

iii. Report regression with mean-centered interaction term

```
##
## Call:
## lm(formula = log.mpg.mc ~ log.acc.mc + log.weight.mc + log.acc.mc *
##     log.weight.mc)
##
## 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)               0.005447   0.008857   0.615 0.538884
## log.acc.mc                0.187500   0.051862   3.615 0.000339 ***
## log.weight.mc            -0.997466   0.031930 -31.239  < 2e-16 ***
## log.acc.mc:log.weight.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
```

```
wei_x_acc <- cars_log$log.weight.*cars_log$log.acceleration.
inter_regr <- lm(wei_x_acc~ cars_log$log.weight.+cars_log$log.acceleration.)
inter_orth <- inter_regr$residuals
summary(lm(log.mpg.~ log.weight.+log.acceleration.+inter_orth,data = cars_log))
```

iv. Report regression with an orthogonalized interaction term

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + inter_orth,
##     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 ***
## log.weight.       -1.00048    0.03187 -31.395  < 2e-16 ***
## log.acceleration.  0.21084    0.04949   4.260 2.56e-05 ***
## inter_orth         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 interaction above, what is the coorelation between interaction and multiplied variables?

```
raw1 <- cor(cars_log$log.weight.*cars_log$log.acceleration.,cars_log$log.weight.)
raw2 <- cor(cars_log$log.weight.*cars_log$log.acceleration.,cars_log$log.acceleration.)
```

raw

```
mean1 <- cor(log.acc.mc*log.weight.mc,log.weight.mc)[1,]
mean2 <- cor(log.acc.mc*log.weight.mc,log.acc.mc)[1,]
```

mean-centered

```
orth1 <- cor(inter_orth, cars_log$log.weight.)
orth2 <- cor(inter_orth,cars_log$log.acceleration.)
```

```
mat <- matrix(c(raw1,raw2,mean1,mean2,orth1,orth2),ncol = 2,byrow = TRUE)
rownames(mat) <- c("raw","mean-centered","orthogonalized")
colnames(mat) <- c("log.weight","log.acceleration")
round(mat,3)
```

orthogonalized

```
##                 log.weight log.acceleration
## raw                  0.108            0.853
## mean-centered       -0.203            0.351
## orthogonalized       0.000            0.000
```