BACS HW13

106070020

2021年5月21日

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

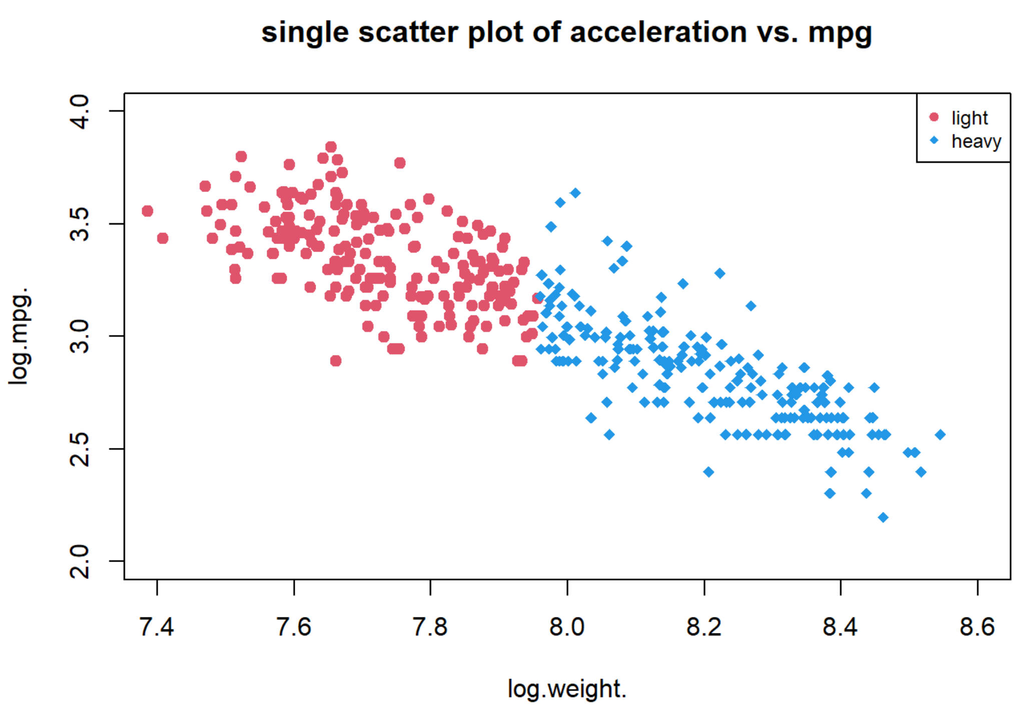
(a) Let’s visualize how weight might moderate the relationship between acceleration and mpg:

(i)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)

```
cars <- read.table("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/BACS/HW11 BACS/auto-data.txt", header=F, 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(cylinders), log(displacement), log(horsepower), log(weight), log(accel
eration), model_year, origin))
lmw<-mean(cars_log$log.weight.)
light <- subset(cars_log, log.weight.<lmw, na.action=na.exclude)
heavy <- subset(cars_log, log.weight.>lmw, na.action=na.exclude)
```

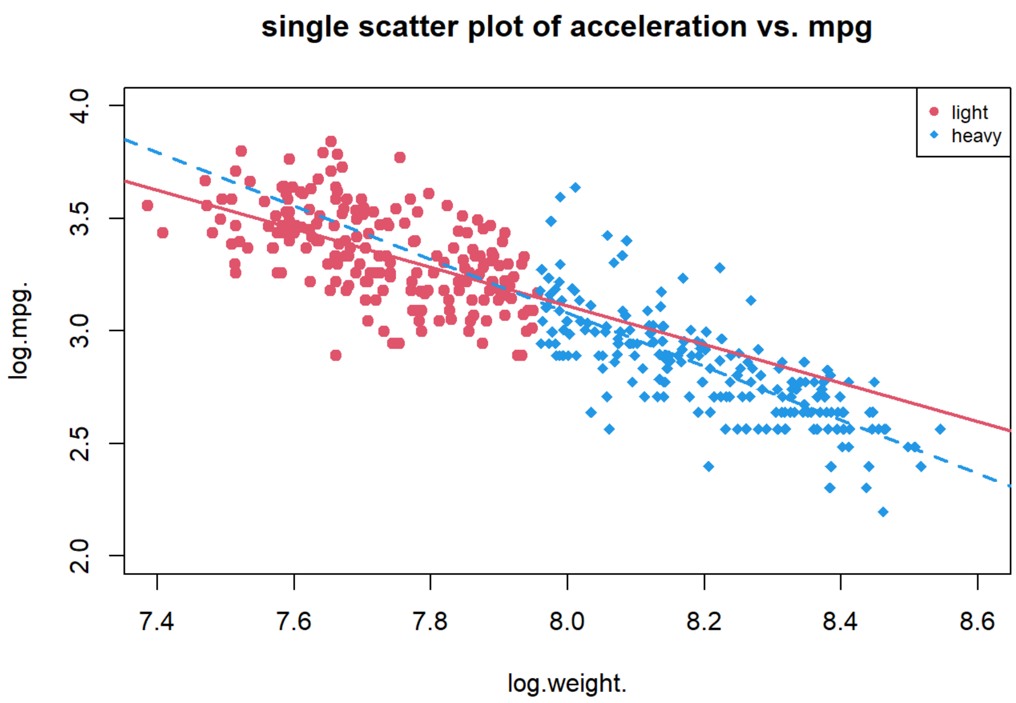
(ii)Create a single scatter plot of acceleration vs. mpg, with different colors and/or shapes for light versus heavy cars

```
#(ii)
with(light, plot(log.weight., log.mpg.,xlim=c(7.4,8.6),ylim=c(2,4), pch=19, col=2, main='single scatter plot of accele
ration vs. mpg'))
with(heavy, points(log.weight., log.mpg., pch=18, col=4))
legend('topright', legend=c("light", "heavy"),
       col=c(2,4), pch=c(19,18), cex=0.75)
```



(iii)Draw two slopes of acceleration-vs-mpg over the scatter plot: one slope for light cars and one slope for heavy cars (use different line styles)

```
#(iii)
with(light, plot(log.weight., log.mpg.,xlim=c(7.4,8.6),ylim=c(2,4), pch=19, col=2, main='single scatter plot of accele
ration vs. mpg'))
with(heavy, points(log.weight., log.mpg., pch=18, col=4))
abline(lm(log.mpg. ~ log.weight., data=light), col=2, lwd=2)
abline(lm(log.mpg. ~ log.weight., data=heavy), col=4, lwd=2, lty=2)
legend('topright', legend=c("light", "heavy"),
       col=c(2,4), pch=c(19,18), cex=0.75)
```



(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

```
light<-na.omit(light)
heavy<-na.omit(heavy)
l<-lm(light$log.mpg.~light$log.weight.+light$log.acceleration.+light$model_year+factor(light$origin))
summary(l)
```

```
##
## Call:
## lm(formula = light$log.mpg. ~ light$log.weight. + light$log.acceleration. +
##     light$model_year + factor(light$origin))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36684 -0.06688  0.00620  0.06448  0.31576
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              6.817512   0.606080  11.249   <2e-16 ***
## light$log.weight.       -0.820783   0.066717 -12.302   <2e-16 ***
## light$log.acceleration.  0.111434   0.058800   1.895   0.0595 .
## light$model_year         0.033109   0.002096  15.798   <2e-16 ***
## factor(light$origin)2    0.039695   0.021455   1.850   0.0658 .
## factor(light$origin)3    0.020798   0.019458   1.069   0.2864
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1109 on 196 degrees of freedom
## Multiple R-squared:  0.7034, Adjusted R-squared:  0.6958
## F-statistic: 92.97 on 5 and 196 DF,  p-value: < 2.2e-16
```

```
h<-lm(heavy$log.mpg.~heavy$log.weight.+heavy$log.acceleration.+heavy$model_year+factor(heavy$origin))
summary(h)
```

```
##
## Call:
## lm(formula = heavy$log.mpg. ~ heavy$log.weight. + heavy$log.acceleration. +
##     heavy$model_year + factor(heavy$origin))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37106 -0.07150  0.00276  0.06702  0.42505
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              7.096619   0.690120  10.283  < 2e-16 ***
## heavy$log.weight.       -0.824266   0.069657 -11.833  < 2e-16 ***
## heavy$log.acceleration.  0.031170   0.056250   0.554  0.58017
## heavy$model_year         0.032086   0.003325   9.649  < 2e-16 ***
## factor(heavy$origin)2    0.098291   0.034250   2.870  0.00459 **
## factor(heavy$origin)3    0.061596   0.066222   0.930  0.35351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.122 on 184 degrees of freedom
## Multiple R-squared:  0.754,  Adjusted R-squared:  0.7473
## F-statistic: 112.8 on 5 and 184 DF,  p-value: < 2.2e-16
```

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

> Both log.weight. and model_year have the significant effects on both light and heavy cars. While the log.acceleration. is only have a significant effect on light cars at 10% significance.

Question 2

(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.

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

```
##
## 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
```

> In my opinion, log.acceleration. 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

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

```
##
## 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
```

(ii)Report a regression with an interaction between weight and acceleration

```
cars_log2<-with(cars_log, data.frame(log.mpg.,  log.weight.,  log.acceleration., model_year, origin))
regr_1<- lm(log.mpg. ~ log.weight. + log.acceleration. + model_year +
              factor(origin)+log.weight.*log.acceleration.,  data=cars_log2)
summary(regr_1)
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin) + log.weight. * log.acceleration., data = cars_log2)
##
## 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
```

```
cor(cbind(cars_log2,cars_log2$log.weight.*cars_log2$log.acceleration.))
```

```
##                                                      log.mpg. log.weight.
## log.mpg.                                          1.000000000  -0.8744686
## log.weight.                                      -0.874468594   1.0000000
## log.acceleration.                                 0.464053310  -0.4256194
## model_year                                        0.576342261  -0.2840090
## origin                                            0.558329285  -0.6048831
## cars_log2$log.weight. * cars_log2$log.acceleration.  0.007445392   0.1083055
##                                                  log.acceleration.
## log.mpg.                                                 0.4640533
## log.weight.                                             -0.4256194
## log.acceleration.                                        1.0000000
## model_year                                               0.3107471
## origin                                                   0.2210906
## cars_log2$log.weight. * cars_log2$log.acceleration.         0.8528810
##                                                  model_year     origin
## log.mpg.                                          0.5763423  0.5583293
## log.weight.                                      -0.2840090 -0.6048831
## log.acceleration.                                 0.3107471  0.2210906
## model_year                                        1.0000000  0.1806622
## origin                                            0.1806622  1.0000000
## cars_log2$log.weight. * cars_log2$log.acceleration.  0.1853457 -0.1078488
##                                                  cars_log2$log.weight. * cars_log2$log.acceleration.
## log.mpg.                                                                                0.007445392
## log.weight.                                                                             0.108305532
## log.acceleration.                                                                       0.852881042
## model_year                                                                              0.185345672
## origin                                                                                 -0.107848822
## cars_log2$log.weight. * cars_log2$log.acceleration.                                     1.000000000
```

(iii)Report a regression with an interaction between weight and acceleration

```
mlogw<-scale(cars_log2$log.weight.)
mloga<-scale(cars_log2$log.acceleration.)
regr_2<-lm(cars_log2$log.mpg. ~ mlogw + mloga +cars_log2$model_year +
             factor(cars_log2$origin)+ mlogw*mloga)
summary(regr_2)
```

```
##
## Call:
## lm(formula = cars_log2$log.mpg. ~ mlogw + mloga + cars_log2$model_year +
##     factor(cars_log2$origin) + mlogw * mloga)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37807 -0.06868  0.00463  0.06891  0.39857
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.518882   0.132944   3.903 0.000112 ***
## mlogw                     -0.247095   0.008023 -30.799  < 2e-16 ***
## mloga                      0.013120   0.006789   1.932 0.054031 .
## cars_log2$model_year       0.033685   0.001735  19.411  < 2e-16 ***
## factor(cars_log2$origin)2  0.058737   0.017789   3.302 0.001049 **
## factor(cars_log2$origin)3  0.028179   0.018266   1.543 0.123704
## mlogw:mloga               -0.014566   0.006283  -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

```
logw_loga<-cars_log2$log.weight.*cars_log2$log.acceleration.
interaction_regr<-lm(logw_loga~cars_log2$log.weight.+cars_log2$log.acceleration.)
interaction_ortho<-interaction_regr$residuals
round(cor(cbind(cars_log2, 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
```

```
summary(lm(log.mpg. ~ log.weight. + log.acceleration. + model_year +
              factor(origin)+interaction_ortho, data=cars_log2))
```

```
##
## Call:
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model_year +
##     factor(origin) + interaction_ortho, data = cars_log2)
##
## 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
## 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
## interaction_ortho -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
```

(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_log2$log.weight., cars_log2$log.weight.*cars_log2$log.acceleration.)
raw
```

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

```
mc<-as.numeric(cor(mlogw, mlogw*mloga))
mc
```

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

```
oth<-round(cor(cbind(cars_log2$log.weight., interaction_ortho)),2)
oth
```

```
##                     interaction_ortho
##                   1                 0
## interaction_ortho 0                 1
```