

HW13

106011138

2021/5/23

Question 1

```
library(ggplot2)
library(GGally)

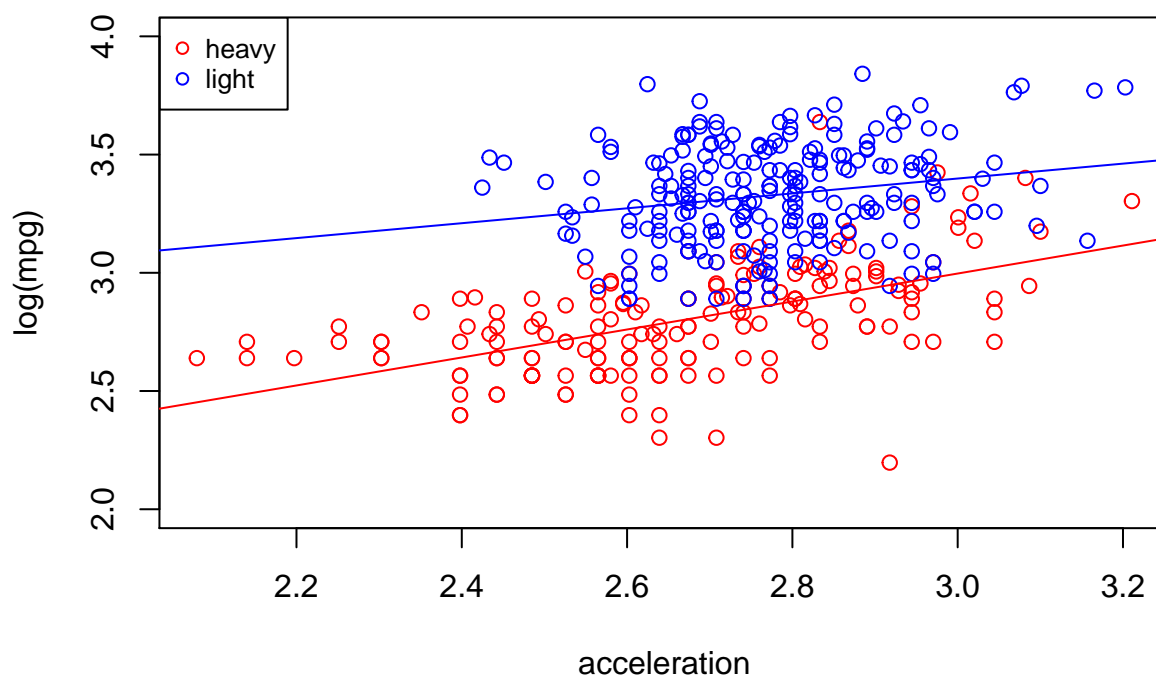
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

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(cylinders), log(displacement),
                                log(horsepower), log(weight), log(acceleration), model_year, origin))
```

a

```
# subset
cars_heavy = subset(cars_log, log.weight.> log(mean(cars$weight)))
cars_light = subset(cars_log, log.weight.< log(mean(cars$weight)))
# scatter plot
plot(cars_heavy$log.acceleration., cars_heavy$log.mpg.,
     ylim = c(2,4), main = 'acceleration vs. mpg',
     xlab = 'acceleration', ylab = 'log(mpg)', col = 'red')
points(cars_light$log.acceleration., cars_light$log.mpg., col = 'blue')
abline(lm(log.mpg.~log.acceleration., data=cars_heavy), col = 'red')
abline(lm(log.mpg.~log.acceleration., data=cars_light), col='blue')
legend("topleft", cex=0.8, pt.cex=0.8, pch=c(1,1), col=c("red", 'blue'),
legend = c('heavy', 'light'))
```

acceleration vs. mpg



b
light cars

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

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

heavy cars

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

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

According to the scatter plot, there is larger mpg in same acceleration for light cars.

Question2

b i

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

```
##
## 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

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

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

```
log_weight_mean = scale(cars_log$log.weight., center=TRUE, scale=FALSE)
log_acceleration_mean = scale(cars_log$log.acceleration., center=TRUE, scale=FALSE)
summary(lm(log.mpg.~log_weight_mean+log_acceleration_mean+log_weight_mean*log_acceleration_mean+model_y

##
```

```
## Call:
## lm(formula = log.mpg. ~ log_weight_mean + log_acceleration_mean +
##     log_weight_mean * log_acceleration_mean + model_year + factor(origin),
##     data = cars_log)
##
## 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_mean    -0.880393   0.028585 -30.799 < 2e-16 ***
## log_acceleration_mean  0.072596   0.037567   1.932 0.054031 .
## model_year         0.033685   0.001735  19.411 < 2e-16 ***
## factor(origin)2     0.058737   0.017789   3.302 0.001049 **
## factor(origin)3     0.028179   0.018266   1.543 0.123704
## log_weight_mean:log_acceleration_mean -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

```
weight_x_acc = cars_log$log.weight.*cars_log$log.acceleration.
interaction_regr = lm(weight_x_acc ~cars_log$log.weight.+cars_log$log.acceleration.)
interaction_ortho = interaction_regr$residuals
summary(lm(log.mpg.~log.weight.+log.acceleration.+interaction_ortho+model_year+factor(origin),data=cars_log))
```

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

```
round(cor(cbind(cars_log['log.weight.'], cars_log['log.acceleration.'], weight_x_acc)), 2)
```

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

```
residuals_iii = cbind(log_weight_mean, log_acceleration_mean,
log_weight_mean*log_acceleration_mean)
colnames(residuals_iii) = c('log_weight_mean', 'log_acc_mean',
'log_weight_mean:log_acc_mean')
round(cor(residuals_iii), 2)
```

```
##           log_weight_mean log_acc_mean
## log_weight_mean           1.00      -0.43
## log_acc_mean              -0.43       1.00
## log_weight_mean:log_acc_mean -0.20       0.35
##           log_weight_mean:log_acc_mean
## log_weight_mean              -0.20
## log_acc_mean                 0.35
## log_weight_mean:log_acc_mean           1.00
```

```
round(cor(cbind(cars_log['log.weight.'], cars_log['log.acceleration.'],
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
```

Credit Help:

106011153

106011204