bacs\_hw10

110071010

2024-04-27

# Pre-setting

Create a data.frame called cars\_log with log-transformed columns for mpg, weight, and acceleration (model\_year and origin don’t have to be transformed)

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, origin))

# Question 1

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

## 1a

Let’s visualize how weight might moderate the relationship between acceleration and mpg

### i

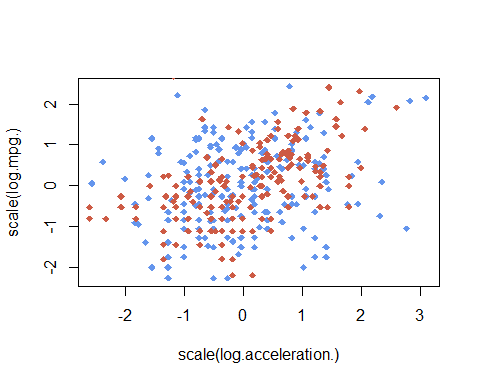
Create two subsets of your data, one for light-weight cars (less than mean weight) and one for heavy cars (higher than the mean weight)

mm <- mean(cars\_log$log.weight.)  
cars\_light <- subset(cars\_log, log.weight. < mm)  
cars\_heavy <- subset(cars\_log, log.weight. > mm)

### ii

Create a single scatterplot of acceleration vs. mpg, with different colors and/or shapes for light versus heavy cars

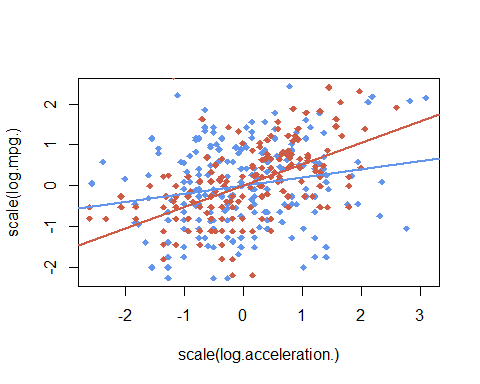
weight\_colors = c("cornflowerblue", "coral3")  
with(cars\_light, plot(scale(log.acceleration.), pch = 18, scale(log.mpg.), col=weight\_colors[1]))  
with(cars\_heavy, points(scale(log.acceleration.), pch = 18, scale(log.mpg.), col=weight\_colors[2]))



### iii

Draw two slopes of acceleration-vs-mpg over the scatter plot: one slope for light cars and one slope for heavy cars (distinguish them by appearance)

weight\_colors = c("cornflowerblue", "coral3")  
with(cars\_light, plot(scale(log.acceleration.), pch = 18, scale(log.mpg.), col=weight\_colors[1]))  
with(cars\_heavy, points(scale(log.acceleration.), pch = 18, scale(log.mpg.), col=weight\_colors[2]))  
  
with(cars\_light, abline(lm(scale(log.acceleration.)~scale(log.mpg.)), col=weight\_colors[1], lwd=2))  
with(cars\_heavy, abline(lm(scale(log.acceleration.)~scale(log.mpg.)), col=weight\_colors[2], lwd=2))



## 1b

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 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   
## -0.36590 -0.06612 0.00637 0.06333 0.31513   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.809014 0.598446 11.378 <2e-16 \*\*\*  
## log.weight. -0.821951 0.065769 -12.497 <2e-16 \*\*\*  
## log.acceleration. 0.111137 0.058297 1.906 0.0580 .   
## model\_year 0.033344 0.002049 16.270 <2e-16 \*\*\*  
## factor(origin)2 0.042309 0.020926 2.022 0.0445 \*   
## factor(origin)3 0.020923 0.019210 1.089 0.2774   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1102 on 199 degrees of freedom  
## Multiple R-squared: 0.7093, Adjusted R-squared: 0.702   
## F-statistic: 97.1 on 5 and 199 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   
## -0.37099 -0.07224 0.00150 0.06704 0.42751   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.132892 0.677740 10.525 < 2e-16 \*\*\*  
## log.weight. -0.825517 0.068101 -12.122 < 2e-16 \*\*\*  
## log.acceleration. 0.031221 0.055465 0.563 0.57418   
## model\_year 0.031735 0.003254 9.752 < 2e-16 \*\*\*  
## factor(origin)2 0.099027 0.033840 2.926 0.00386 \*\*   
## factor(origin)3 0.063148 0.065535 0.964 0.33650   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1212 on 187 degrees of freedom  
## Multiple R-squared: 0.7585, Adjusted R-squared: 0.752   
## F-statistic: 117.4 on 5 and 187 DF, p-value: < 2.2e-16

## 1c

**Question**

Using your intuition only: What do you observe about light versus heavy cars so far?

**Answer**

Heavy cars display stronger acceleration-mpg relationship.

# Question 2

Use the transformed dataset from above (cars\_log), to test whether we have moderation.

## 2a

**Instruction**

Considering weight and acceleration, use your intuition and experience to state which of the two variables might be a moderating versus independent variable, in affecting mileage.

**Answer**

Weight is the moderator,and acceleration is the independent variable.

## 2b

Use various regression models to model the possible moderation on log.mpg.:  
(use log.weight., log.acceleration., model\_year and origin as independent variables)

### i

Report a regression without any interaction terms

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)  
##   
## Coefficients:  
## (Intercept) log.weight. log.acceleration. model\_year   
## 7.43116 -0.87661 0.05151 0.03273   
## factor(origin)2 factor(origin)3   
## 0.05799 0.03233

### ii

Report a regression with an interaction between weight and acceleration

lm(log.mpg. ~ log.weight.+log.acceleration.+model\_year+factor(origin)+log.weight.\* log.acceleration., data=cars\_log)

##   
## Call:  
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model\_year +   
## factor(origin) + log.weight. \* log.acceleration., data = cars\_log)  
##   
## Coefficients:  
## (Intercept) log.weight.   
## 1.08964 -0.09663   
## log.acceleration. model\_year   
## 2.35757 0.03368   
## factor(origin)2 factor(origin)3   
## 0.05874 0.02818   
## log.weight.:log.acceleration.   
## -0.28717

### iii

Report a regression with a mean-centered interaction term

log.weight.\_mean\_centered <- scale(cars\_log$log.weight., center=TRUE, scale=FALSE)  
log.acceleration.\_mean\_centered <- scale(cars\_log$log.acceleration., center=TRUE, scale=FALSE)

lm(log.mpg. ~ log.weight.\_mean\_centered +log.acceleration.\_mean\_centered +model\_year+factor(origin)+log.weight.\_mean\_centered \*log.acceleration.\_mean\_centered , data=cars\_log)

##   
## Call:  
## lm(formula = log.mpg. ~ log.weight.\_mean\_centered + log.acceleration.\_mean\_centered +   
## model\_year + factor(origin) + log.weight.\_mean\_centered \*   
## log.acceleration.\_mean\_centered, data = cars\_log)  
##   
## Coefficients:  
## (Intercept)   
## 0.51888   
## log.weight.\_mean\_centered   
## -0.88039   
## log.acceleration.\_mean\_centered   
## 0.07260   
## model\_year   
## 0.03368   
## factor(origin)2   
## 0.05874   
## factor(origin)3   
## 0.02818   
## log.weight.\_mean\_centered:log.acceleration.\_mean\_centered   
## -0.28717

### iv

Report a regression with an orthogonalized interaction term

interaction\_regr <- lm(cars\_log$log.weight. \* cars\_log$log.acceleration. ~ cars\_log$log.weight.+cars\_log$log.acceleration.)  
interaction\_ortho <- interaction\_regr$residuals  
lm(log.mpg. ~ log.weight.+log.acceleration.+model\_year+factor(origin)+ interaction\_ortho, data=cars\_log)

##   
## Call:  
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model\_year +   
## factor(origin) + interaction\_ortho, data = cars\_log)  
##   
## Coefficients:  
## (Intercept) log.weight. log.acceleration. model\_year   
## 7.37718 -0.87697 0.04610 0.03368   
## factor(origin)2 factor(origin)3 interaction\_ortho   
## 0.05874 0.02818 -0.28717

## 2c

**Question**

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?

**Answer**

interaction\_term – log.weight. : 0.108305532

interaction\_term – log.acceleration. : 0.852881042

cor(cbind(interaction\_term = cars\_log$log.weight \* cars\_log$log.acceleration.,cars\_log))

## interaction\_term log.mpg. log.weight. log.acceleration.  
## interaction\_term 1.000000000 0.007445392 0.1083055 0.8528810  
## log.mpg. 0.007445392 1.000000000 -0.8744686 0.4640533  
## log.weight. 0.108305532 -0.874468594 1.0000000 -0.4256194  
## log.acceleration. 0.852881042 0.464053310 -0.4256194 1.0000000  
## model\_year 0.185345672 0.576342261 -0.2840090 0.3107471  
## origin -0.107848822 0.558329285 -0.6048831 0.2210906  
## model\_year origin  
## interaction\_term 0.1853457 -0.1078488  
## 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

# Question 3

Let’s check whether weight mediates the relationship between cylinders and mpg, even when other factors are controlled for.  Use log.mpg., log.weight., and log.cylinders as your main variables, and keep log.acceleration., model\_year, and origin as control variables.

## 3a

Let’s try computing the direct effects first:

### i

**Instruction**

Model 1: Regress log.weight. over log.cylinders. only

Check whether number of cylinders has a significant direct effect on weight

**Observation**

Yes. The coefficient of 0.8201 reflects a strong positive relationship.

cars\_log1 <- with(cars, data.frame(log(mpg), log(weight), log(acceleration),log(cylinders), model\_year, origin))  
A <- lm(log.weight. ~ log.cylinders., data=cars\_log1)  
A

##   
## Call:  
## lm(formula = log.weight. ~ log.cylinders., data = cars\_log1)  
##   
## Coefficients:  
## (Intercept) log.cylinders.   
## 6.6037 0.8201

### ii

**Instruction**

Model 2: Regress log.mpg. over log.weight. and all control variables

Check whether weight has a significant direct effect on mpg with other variables statistically controlled

**Observation**

Yes. The coefficient of -0.87661 reflects a strong negative relationship.

B <- lm(log.mpg. ~ log.weight.+log.acceleration.+model\_year+factor(origin), data=cars\_log1)  
B

##   
## Call:  
## lm(formula = log.mpg. ~ log.weight. + log.acceleration. + model\_year +   
## factor(origin), data = cars\_log1)  
##   
## Coefficients:  
## (Intercept) log.weight. log.acceleration. model\_year   
## 7.43116 -0.87661 0.05151 0.03273   
## factor(origin)2 factor(origin)3   
## 0.05799 0.03233

## 3b

What is the indirect effect of cylinders on mpg? (use the product of slopes between Models 1 & 2)

(A$coefficients[2]) \* (B$coefficients[2])

## log.cylinders.   
## -0.7189275

## 3c

Let’s bootstrap for the confidence interval of the indirect effect of cylinders on mpg

### i

Bootstrap regression models 1 & 2, and compute the indirect effect each time:  
What is its 95% CI of the indirect effect of log.cylinders. on log.mpg.?

boot\_mediation <- function(model\_a, model\_b, dataset) {  
 boot\_index <- sample(1:nrow(dataset), replace=TRUE)  
 data\_boot <- dataset[boot\_index, ]  
 regr1 <- lm(model\_a, data\_boot)  
 regr2 <- lm(model\_b, data\_boot)  
 return(regr1$coefficients[2] \* regr2$coefficients[2])  
}  
  
set.seed(12341234)  
indirect <- replicate(2000, boot\_mediation(A, B, cars\_log1))  
quantile(indirect, probs=c(0.025, 0.975))

## 2.5% 97.5%   
## -0.7823354 -0.6604392

### ii

Show a density plot of the distribution of the indirect effect, and mark its 95% CI

plot(density(indirect),main = "Indirect Effect Distribution", col = "cornflowerblue", lwd = 2)  
abline(v = quantile(indirect, probs=c(0.025, 0.975)), col = "gray", lwd = 2)

