

HW2

Peiran Chen

4/13/2022

1.

```
head(Auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307         130   3504          12.0    70     1
## 2  15         8          350         165   3693          11.5    70     1
## 3  18         8          318         150   3436          11.0    70     1
## 4  16         8          304         150   3433          12.0    70     1
## 5  17         8          302         140   3449          10.5    70     1
## 6  15         8          429         198   4341          10.0    70     1
##                                name
## 1 chevrolet chevelle malibu
## 2      buick skylark 320
## 3    plymouth satellite
## 4          amc rebel sst
## 5                ford torino
## 6          ford galaxie 500
```

```
lm_1 <- lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration + year + origin, data = Auto)
summary(lm_1)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.5903 -2.1565 -0.1169  1.8690 13.0604
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
## cylinders      -0.493376   0.323282  -1.526  0.12780
## displacement   0.019896   0.007515   2.647  0.00844 **
## horsepower     -0.016951   0.013787  -1.230  0.21963
## weight         -0.006474   0.000652  -9.929 < 2e-16 ***
## acceleration   0.080576   0.098845   0.815  0.41548
```

```
## year          0.750773    0.050973   14.729   < 2e-16 ***
## origin        1.426141    0.278136    5.127 4.67e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.328 on 384 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8182
## F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

Yes, there is a relationship between the predictors and the response by testing the null hypothesis of whether all the regression coefficients are zero. The F-statistic is far from 1 (with a small p-value), indicating evidence against the null hypothesis.

b.

```
train_MSE <- mean(lm_1$residuals^2)
```

The train MSE in this linear model is 10.8474809.

c.

Since it's not hard to see that Origin = 3 means a Japanese car

```
# Filter Data into only Japanese Car
# Auto_Japan <- Auto %>%
#   filter(origin == 3)

# This time, I choose to drop Year and fit a new model
# lm_2 <- lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration, data = Auto_Japan)

prediction_1 <- predict(lm_1, data.frame(cylinders = 3, displacement = 100, horsepower = 85, weight = 3000))
```

The mileage my model predict for the given car is 1454.8478639