HW 1 Steven(sap407) and Piyaporn (pp712)

Class: 33:136:487:01 LG SCALE DATA ANALY

Prof. Jin Wang

Piyaporn Puangprasert(pp712) and Steven Panagakos(sap407)

Due date: Feb 15, 2024

——————————————————————————

read the data file

```
dat = read.csv("Auto1a.csv")
names(dat)
```

```
## [1] "mpg"          "cylinders"    "displacement" "horsepower"   "weight"
## [6] "acceleration" "year"         "origin"       "name"
```

model-1: mpg~horsepower

```
model = lm(mpg~horsepower, data = dat)
```

coefficients

```
model$coefficients
```

```
## (Intercept)   horsepower
##  38.4121484  -0.1484225
```

CI

```
confint.lm(model)
```

```
##                  2.5 %      97.5 %
## (Intercept) 36.9704297 39.8538670
## horsepower  -0.1610343 -0.1358106
```

summary

```
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ horsepower, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.726  -3.149  -0.633   2.532  17.835
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.412148   0.733004   52.40   <2e-16 ***
## horsepower  -0.148422   0.006412  -23.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.74 on 345 degrees of freedom
## Multiple R-squared:  0.6083, Adjusted R-squared:  0.6072
## F-statistic: 535.8 on 1 and 345 DF,  p-value: < 2.2e-16
```

I. Is there a relationship between the predictor and the response? Why?

Answer:

II. How strong is the relationship between the predictor and the response?

Answer:

III. Is the relationship between the predictor and the response positive or negative?

Answer:

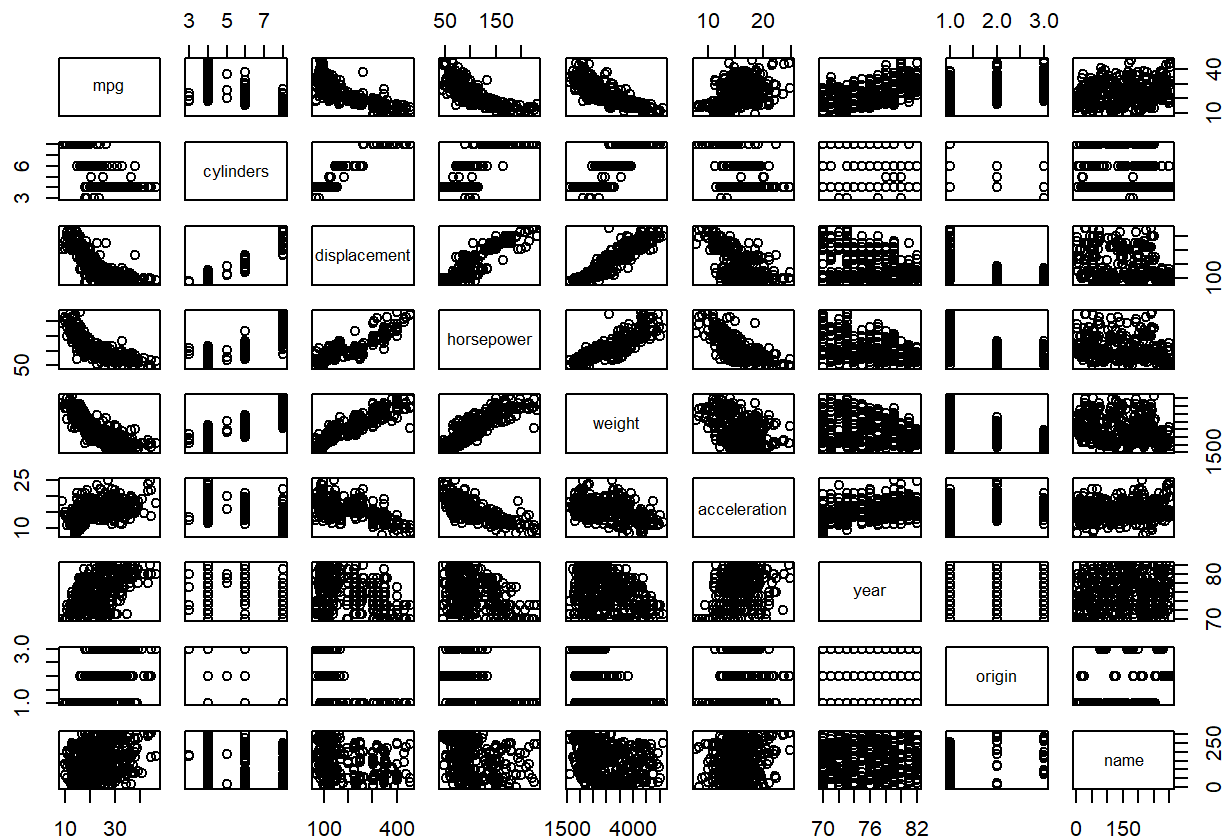IV What is the Predicted mpg with hoursepower of 98%

```
predict_mpg = predict(model, data.frame(horsepower = 98))
predict_mpg
```

```
##        1
## 23.86675
```

2. mutiple linear regression

Scatterplot matrix that include all the variable in the dataset

```
library(ISLR2)
dat = read.csv("Auto1a.csv")
pairs(Auto)
```

Compute the matrix of correlations between the variables using the function

```
cor= cor(dat[,-9])
cor
```

```
##                    mpg  cylinders displacement horsepower     weight
## mpg          1.0000000 -0.7830225   -0.8151730 -0.7799376 -0.8407863
## cylinders   -0.7830225  1.0000000    0.9518758  0.8442297  0.8999313
## displacement -0.8151730  0.9518758    1.0000000  0.9000725  0.9342206
## horsepower  -0.7799376  0.8442297    0.9000725  1.0000000  0.8652063
## weight      -0.8407863  0.8999313    0.9342206  0.8652063  1.0000000
## acceleration 0.4316216 -0.5243642   -0.5644714 -0.7053185 -0.4375583
## year         0.5153320 -0.2844881   -0.3237214 -0.3789386 -0.2588493
## origin       0.5980460 -0.6095630   -0.6445736 -0.4802563 -0.6091412
##              acceleration       year     origin
## mpg             0.4316216  0.5153320  0.5980460
## cylinders      -0.5243642 -0.2844881 -0.6095630
## displacement   -0.5644714 -0.3237214 -0.6445736
## horsepower     -0.7053185 -0.3789386 -0.4802563
## weight         -0.4375583 -0.2588493 -0.6091412
## acceleration    1.0000000  0.2751593  0.2566660
## year            0.2751593  1.0000000  0.1762578
## origin          0.2566660  0.1762578  1.0000000
```

Use the lm() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summary() function to print the results. Comment

on the output. For instance:

Is there a relationship between the predictors and the response?

Which predictors appear to have a statistically significant relationship to the response?

What does the coefficient for the year variable suggest?

```
model4 = lm(mpg~. - name, data = dat)
summary(model4)
```

```
##
## Call:
## lm(formula = mpg ~ . - name, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.8040 -2.0424 -0.1774  1.7970 13.2491
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.568e+01  5.237e+00  -2.995  0.00295 **
## cylinders    -4.047e-01  3.370e-01  -1.201  0.23057
## displacement  1.484e-02  7.857e-03   1.888  0.05984 .
## horsepower   -1.581e-02  1.439e-02  -1.098  0.27277
## weight       -6.099e-03  6.719e-04  -9.077  < 2e-16 ***
## acceleration  2.666e-02  1.055e-01   0.253  0.80076
## year          7.309e-01  6.009e-02  12.164  < 2e-16 ***
## origin        1.441e+00  3.104e-01   4.644  4.9e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.291 on 339 degrees of freedom
## Multiple R-squared:  0.8144, Adjusted R-squared:  0.8106
## F-statistic: 212.5 on 7 and 339 DF,  p-value: < 2.2e-16
```

Use the * symbols to fit linear regression models with interaction effects. Do any interactions appear to be statistically significant?

```
model5 = lm(mpg ~. * . -name,data = dat)

model5
```

```
##
## Call:
## lm(formula = mpg ~ . * . - name, data = dat)
##
## Coefficients:
##                                          (Intercept)
##                                            1.306e+04
##                                             cylinders
##                                            -5.876e+03
##                                          displacement
##                                            1.285e+02
##                                            horsepower
##                                            -7.561e+00
##                                                weight
##                                            -3.445e+00
##                                          acceleration
##                                            7.112e+01
##                                                  year
##                                            -2.053e+02
##                                                origin
##                                            6.665e+02
##                              cylinders:displacement
##                                            -4.962e+00
##                                cylinders:horsepower
##                                            8.837e+00
##                                    cylinders:weight
##                                            -1.129e-01
##                              cylinders:acceleration
##                                            6.618e+00
##                                      cylinders:year
##                                            4.455e+01
##                                    cylinders:origin
##                                            1.033e+03
##                     cylinders:nameamc ambassador dpl
##                                            -2.746e+02
##                     cylinders:nameamc ambassador sst
##                                            3.770e+02
##                          cylinders:nameamc concord
##                                            1.029e+04
##                        cylinders:nameamc concord d/l
##                                            1.804e+02
##                        cylinders:nameamc concord dl 6
##                                            2.676e+02
##                          cylinders:nameamc gremlin
##                                            1.826e+03
##                            cylinders:nameamc hornet
##                                            2.233e+03
##              cylinders:nameamc hornet sportabout (sw)
##                                            1.135e+02
##                          cylinders:nameamc matador
##                                            2.220e+03
##                      cylinders:nameamc matador (sw)
```