Car Performance and Efficiency

Issues

Data set consists of numerical characteristics of several cars of displacement, horsepower, weight, acceleration, miles per gallon (mpg) in which row represent to particular car. The mpg metric is used how these variables related to each other.

We address the questions:

1. Is at least one of the predictors useful in predicting the response?
2. Do all the predictors help to explain the response, or is only a subset of the predictors useful?
3. How well does the model fit the data?
4. Given a set of predictor values, what response value should we predict, and how accurate is our prediction?

Findings

From the analysis only acceleration is positively correlated with mpg while the remaining variables displacement, horsepower, weight are negatively correlated to the mpg. Only displacement and weight are more significant factors when compared to horsepower and acceleration based on the respective p-values.

Discussions

The dataset consists 402 samples, we analyse the given dataset and establish the relationship between the five variables. By looking through the correlation matrix we determine the which variables are significant. By considering the P-value we can say that the model is fit for the dataset.

Appendix A: Method

We imported the .csv fie into the R- studio and analysed how the car efficiency and performance depends on the different factors. Mpg which refers to miles per gallon depends on the weight, displacement, horsepower, acceleration (weight refers to overall weight of car, acceleration refers to rate at which car can change its speed).

As there are more than one factors, we consider the multiple linear regression model by considering corelation coefficients, scatter plots between all variables.

Appendix B: Results

Initially we obatained the correlation value for all factors on depending variable by using the pearson method.

```
> cor(auto, method = "pearson")
              displacement horsepower     weight acceleration        mpg
displacement     1.0000000  0.9181735  0.9260776   -0.5892415 -0.8312848
horsepower       0.9181735  1.0000000  0.8483556   -0.7303687 -0.7733972
weight           0.9260776  0.8483556  1.0000000   -0.4490861 -0.8466882
acceleration    -0.5892415 -0.7303687 -0.4490861    1.0000000  0.4352363
mpg             -0.8312848 -0.7733972 -0.8466882    0.4352363  1.0000000
```

By looking the results we can say that displacement, horsepower, weight are negatively correlated to mpg which is dependent variable, acceleration which is positively correlated(0.435) by looking the value we can say that it is not significant.

We also plotted the scatter plots by using the pairs.panels and analysed how they are related to each other (positively or negatively related to each other).

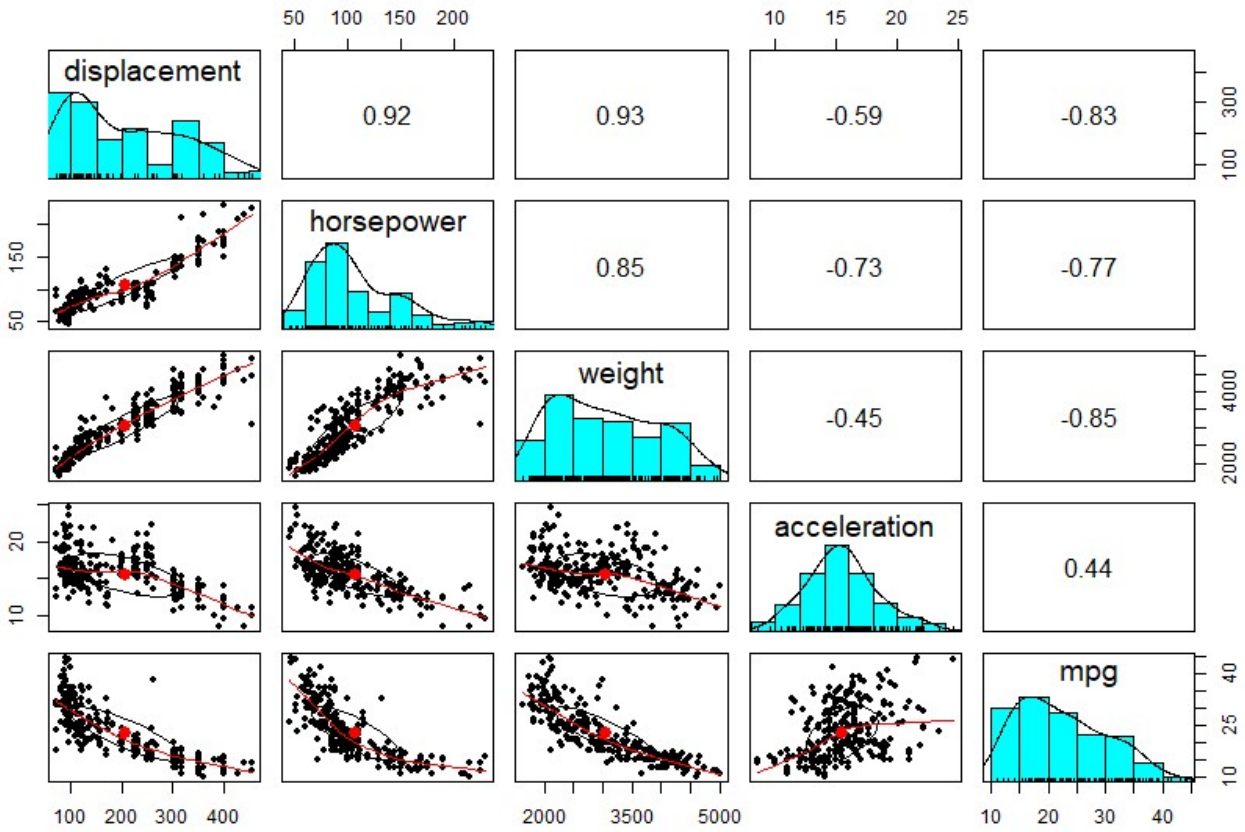

Figure 1. Correlation plots

By seeing the plots we can say that displacement, horsepower and weight are positively related and acceleration is negatively related to the mpg. And coefficient values are also obtained which indicates that how significant to each other.

We used the linear model function to fit the multiple linear regression and obtained results are

```
> m_r<- lm(auto$mpg ~ auto$displacement+auto$horsepower+auto$weight+auto$accelera
tion)
> summary(m_r)

Call:
lm(formula = auto$mpg ~ auto$displacement + auto$horsepower +
    auto$weight + auto$acceleration)

Residuals:
     Min       1Q   Median       3Q      Max
-12.0745  -2.6879  -0.4395   1.8181  15.9470

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       44.7822360  2.3867318  18.763  < 2e-16 ***
auto$displacement -0.0189992  0.0064692  -2.937  0.00351 **
auto$horsepower   -0.0230424  0.0155007  -1.487  0.13793
auto$weight       -0.0045785  0.0006646  -6.889 2.21e-11 ***
auto$acceleration -0.1169871  0.1149550  -1.018  0.30945
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.033 on 397 degrees of freedom
Multiple R-squared:  0.734,    Adjusted R-squared:  0.7313
F-statistic: 273.9 on 4 and 397 DF,  p-value: < 2.2e-16
```

By seeing the summary of the model, we can say that intercept is 44.7822 which is positive value, p- values for displacement and weight are 0.00351, 2.21e-11 respectively which are very less than 0.05 which means they are more significant and it is strong evidence against null hypothesis , p-value for horsepower and acceleration are 0.13793, 0.309 which are greater than 0.05 which means that they are less significant when compared to other two factors. R- squared value of the model is 0.734 which means 73.4% of variation in mpg by changing the independent values. And also, we can obtain the mpg equation how it related to the displacement, weight, acceleration, horsepower.

Mpg = 44.782-(0.01899*displacement)-(0.0230424*horsepower)

-(0.0045785*weight)-(0.1169871*acceleration)

Appendix C: Code

```
install.packages('readxl')

library(readxl)

install.packages("psych")

library(psych)


file <-
"E:\\Assignments\\MTH522\\Project1\\auto_data_muchinapalli_nagendra
.xls"

auto<- read_excel(file,sheet = 1)

summary(auto)

head(auto)


#corelation between variables

cor(auto, method = "pearson")

pairs.panels(auto, cex.cor = 0.5)


#linear model

m_r<- lm(auto$mpg ~
auto$displacement+auto$horsepower+auto$weight+auto$acceleration)

summary(m_r)
```