Multi linear Regression Report on fuel efficiency of Automobile

Issues

Our data collection contains numerical information on a number of vehicle characteristics, such as their weight, engine performance, displacement, acceleration, and miles per gallon.(mpg).

The mpg metric can be used to investigate the relationships between the various vehicle parts and to create a predictive model for a car's fuel efficiency based on a number of different variables.Is at least one of the predictors useful in predicting the response?

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

Findings

According to the analysis, only acceleration has a good relationship with mpg, whereas the other variables—displacement, horsepower, and weight—have a negative relationship with mpg. According to the respective p-values, only weight is a relevant factor when compared to horsepower, acceleration, and displacement.

Discussions

We analyzed the provided dataset and determine the connection between the five variables using the 377 samples that make up the dataset. We identify the important variables by examining the correlation matrix. We can conclude that the model fits the sample by taking the P-value into account.

Appendix A: Method

To investigate the relationship between fuel efficiency and various factors, such as horsepower, acceleration, weight, and displacement, we utilized multiple linear regression analysis due to the presence of multiple predictor variables. We imported the data from a .csv file into R Studio and conducted the analysis. The efficiency of the car was represented by miles per gallon (MPG), while horsepower represented the power of the engine. Acceleration referred to the rate at which the car could change its speed, weight represented the total weight of the car, and displacement measured the cylinder volume swept. We utilized several statistical tools, such as correlation scatter plots, linear model plots, and Pearson's method to analyze the correlation between variables and assess the significance of each variable on the dependent variable.

Appendix B: Results

Correlation value for all factors on depending variable by using the pearson method.

```
> cor(autodata, method = "pearson")
             displacement horsepower    weight acceleration        mpg
displacement    1.0000000  0.8737022  0.9361073   -0.4892218 -0.7843810
horsepower      0.8737022  1.0000000  0.8491263   -0.6663514 -0.7366148
weight          0.9361073  0.8491263  1.0000000   -0.3703465 -0.8176565
acceleration   -0.4892218 -0.6663514 -0.3703465    1.0000000  0.3753230
mpg            -0.7843810 -0.7366148 -0.8176565    0.3753230  1.0000000
```

Based on the analysis results, we can infer that displacement, horsepower, and weight have a negative correlation with the dependent variable, which is miles per gallon (MPG). On the other hand, acceleration has a positive correlation with MPG, although the correlation coefficient (0.375) indicates that it is not significant. To further investigate the relationships between the variables, we created scatter plots using the pairs.panels function and examined whether they were positively or negatively related to each other.
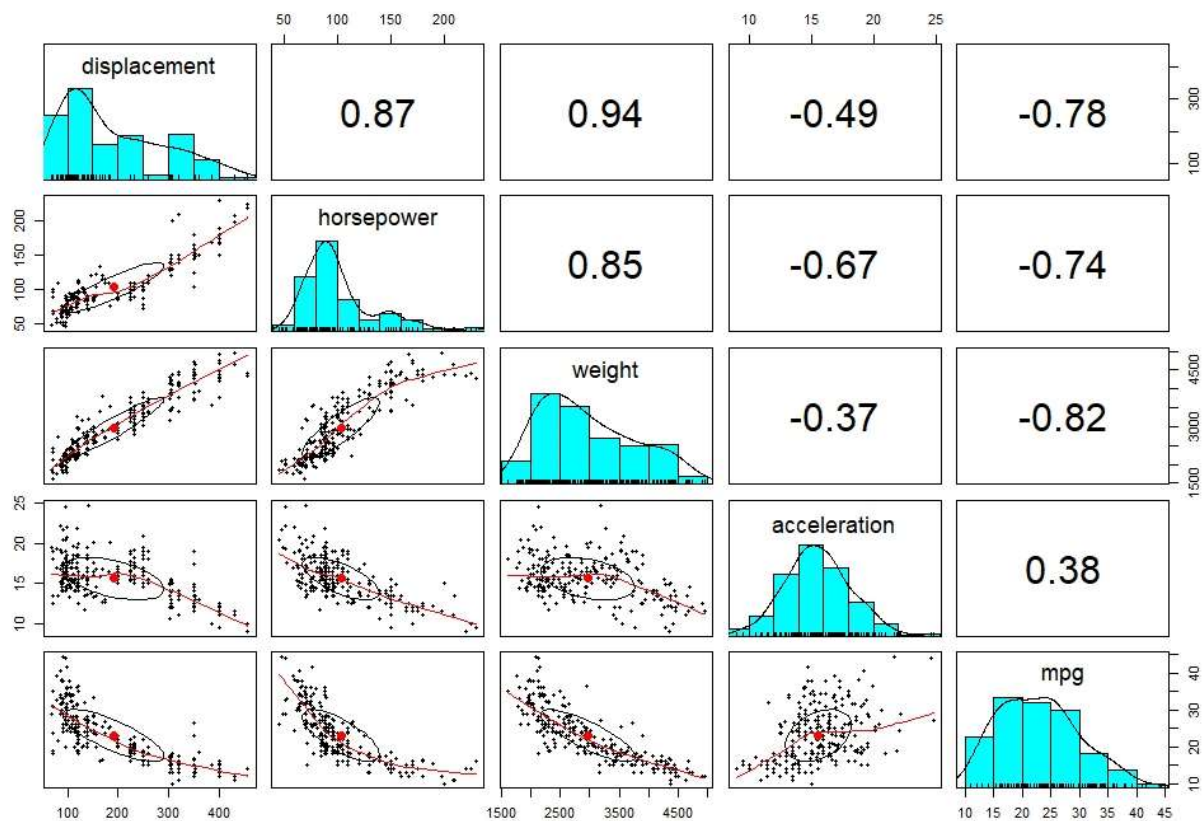
Correlation plots

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

> library(psych)

```
> head(autodata)
  displacement horsepower weight acceleration  mpg
1       120       87   2979      19.5 21.0
2       258      110   3632      18.0 16.0
3       112       88   2395      18.0 34.0
4       225      105   3439      15.5 16.0
5       304      120   3962      13.9 15.5
6       140       89   2755      15.8 25.5
> str(autodata)
Classes 'tbl_df', 'tbl' and 'data.frame':        377 obs. of  5 variables:
 $ displacement: num  120 258 112 225 304 140 91 231 134 85 ...
 $ horsepower  : num  87 110 88 105 120 89 53 110 96 52 ...
 $ weight      : num  2979 3632 2395 3439 3962 ...
 $ acceleration: num  19.5 18 18 15.5 13.9 15.8 17.5 15 13.5 22.2 ...
 $ mpg         : num  21 16 34 16 15.5 25.5 33 21 24 29 ...
> summary(autodata)
  displacement   horsepower      weight      acceleration      mpg
 Min.   : 68.0  Min.   : 46.0  Min.   :1613  Min.   : 9.00  Min.   :10.00
 1st Qu.:110.0  1st Qu.: 79.0  1st Qu.:2278  1st Qu.:14.00  1st Qu.:17.60
 Median :141.0  Median : 92.0  Median :2789  Median :15.50  Median :23.00
 Mean   :190.5  Mean   :103.3  Mean   :2969  Mean   :15.59  Mean   :23.06
 3rd Qu.:250.0  3rd Qu.:110.0  3rd Qu.:3535  3rd Qu.:17.00  3rd Qu.:27.00
 Max.   :455.0  Max.   :230.0  Max.   :4952  Max.   :24.80  Max.   :44.30
> cor(autodata, method = "pearson")
             displacement horsepower     weight acceleration        mpg
displacement    1.0000000  0.8737022  0.9361073   -0.4892218 -0.7843810
horsepower      0.8737022  1.0000000  0.8491263   -0.6663514 -0.7366148
weight          0.9361073  0.8491263  1.0000000   -0.3703465 -0.8176565
acceleration   -0.4892218 -0.6663514 -0.3703465    1.0000000  0.3753230
mpg            -0.7843810 -0.7366148 -0.8176565    0.3753230  1.0000000
> pairs(autodata[1:5])
> pairs.panels(autodata)
> model <- lm(autodata$mpg ~ autodata$displacement+autodata$horsepower+autodata$weight+autodata$acceleration)
> model

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

Coefficients:
      (Intercept) autodata$displacement    autodata$horsepower
        41.697484             -0.003688              -0.016883
    autodata$weight autodata$acceleration
        -0.006012              0.106837

> summary(model)

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

Residuals:
    Min      1Q  Median      3Q     Max
-10.5926 -2.4007 -0.5381  1.9056 13.9616

Coefficients:
```

```
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)          41.6974843  2.4088016  17.310  < 2e-16 ***
autodata$displacement -0.0036878  0.0067232  -0.549   0.584
autodata$horsepower   -0.0168827  0.0159579  -1.058   0.291
autodata$weight      -0.0060121  0.0008578  -7.008 1.14e-11 ***
autodata$acceleration 0.1068368  0.1235718   0.865   0.388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.032 on 372 degrees of freedom
Multiple R-squared:  0.6761,      Adjusted R-squared:  0.6726
F-statistic: 194.1 on 4 and 372 DF,  p-value: < 2.2e-16
```

The linear regression model shows that there is a significant relationship between the dependent variable (mpg) and the independent variables (displacement, horsepower, weight, and acceleration). The intercept value is 41.6974843, which indicates that when all independent variables are zero, the predicted value of the dependent variable is 41.6974843. The coefficients of displacement, horsepower, weight, and acceleration are -0.0036878, -0.0168827, -0.0060121, and 0.1068368, respectively. Only weight has a significant impact on the dependent variable, as evidenced by its very low p-value of 1.14e-11. In contrast, the p-values of the other three independent variables are greater than 0.05, indicating that they are not significant predictors of the dependent variable. The R-squared value of the model is 0.6761, meaning that 67.61% of the variation in the dependent variable can be explained by the independent variables. The adjusted R-squared value of 0.6726 indicates that the model's fit is not significantly improved by adding more independent variables. The F-statistic of 194.1 with a very low p-value indicates that the model is statistically significant overall. The residual standard error is 4.032, indicating that the model has a good fit to the data.

$$Mpg = 41.6974 - (0.003687 * displacement) - (0.0168827 * horsepower)$$
$$- (0.0060121 * weight) + (0.1068368 * acceleration)$$

Appendix C: Code

```
library(psych)
head(autodata)
str(autodata)
summary(autodata)
cor(autodata, method = "pearson")
pairs(autodata[1:5])
pairs.panels(autodata)
model <- lm(autodata$mpg ~
autodata$displacement+autodata$horsepower+autodata$weight+autodata$a
cceleration)
model
summary(model)
```