Trends of Cars using MPG

AMAN

October 26, 2019

Made using Knitr

Synopsis

This report analizes the relationship between the transmission type used in a vehicle and miles per gallon of milage. The report set out to determine which transmission type gives us a higher milage. The vehicles under study are from the mtcars dataset. A t-test between automatic and manual transmission vehicles shows that manual transmission vehicles have a greater MPG than automatic transmission vehicles by approx 7.25 MPG. After fitting multiple linear regression lines, analysis showed that the manual transmission contributed less significantly to MPG, only an improvement of 1.81 MPG. Other variables, weight, horsepower, and number of cylinders contributed more significantly to the overall MPG of vehicles.

Load Data

Load the mtcars dataset and convert variables to factors.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
data(mtcars)
head(mtcars, n=3)
dim(mtcars)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$am <- factor(mtcars$am)
mtcars$vs <- as.factor(mtcars$vs)
attach(mtcars)
```

Exploratory Analysis

The Box graph that compares the milage of Automatic and Manual transmission. The graph leads us to believe that there is a significant increase in MPG when for vehicles with a manual transmission vs automatic.

Statistical Inference

T-Test

```
testResults <- t.test(mpg ~ am)
testResults$p.value
```

```
## [1] 0.001373638
```

The T-Test rejects the null hypothesis that the difference between transmission types is 0.

```
testResults$estimate
```

```
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

The difference estimate between the 2 transmissions is 7.24494 MPG in favor of manual.

Regression

Fit the full model of the available data

Since none of the coefficients have a p-value less than 0.05 we cannot conclude which variables are more statistically significant.

```
step_Fit <- step(full_Model_Fit)
summary(step_Fit)
summary(step_Fit)$coeff
```

The new model made has four variables (horsepower, cylinders, transmission, weight). The R-squared value of 0.8659 tells us that this model explains about 87% of the variance in MPG. The p-values also are statistically significantly because they have a p-value less than 0.05. The coefficients conclude that increasing the number of cylinders from 4 to 6 with decrease the MPG by 3.03.

Residuals & Diagnostics

Residual Plot

The plots conclude:

1. The randomness of the Residuals vs. Fitted plot supports the assumption of independence
2. The points of the Normal Q-Q plot following closely to the line conclude that the distribution of residuals is normal
3. The Scale-Location plot random distribution confirms the constant variance assumption
4. Since all points are within the 0.05 lines, the Residuals vs. Leverage concludes that there are no outliers

```
sum((abs(dfbetas(step_Fit)))>1)
```

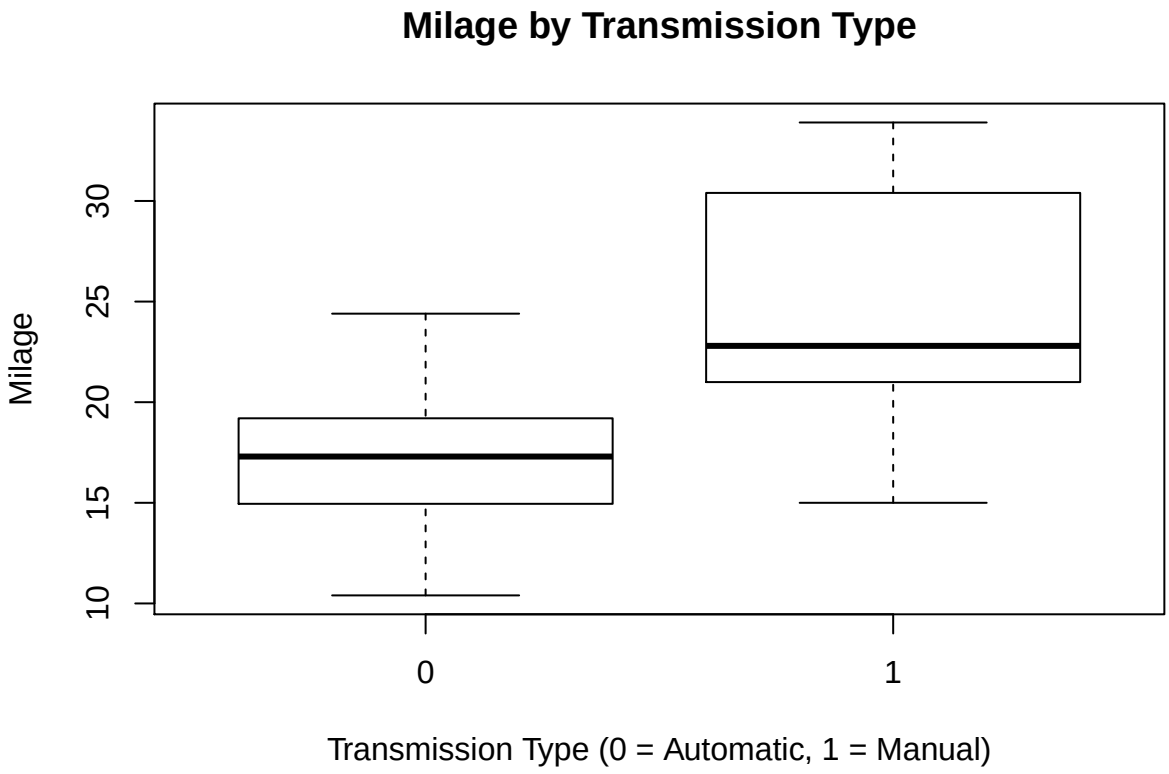
```
## [1] 0
```

Conclusion

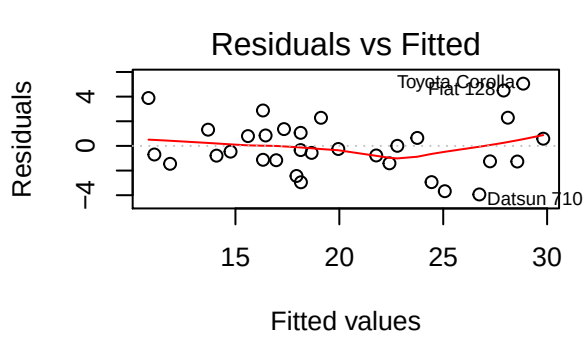
There is a difference in MPG based on transmission type. A manual transmission will have a slight MPG boost. However, it seems that weight, horsepower, & number of cylinders are more statistically significant when determining MPG.

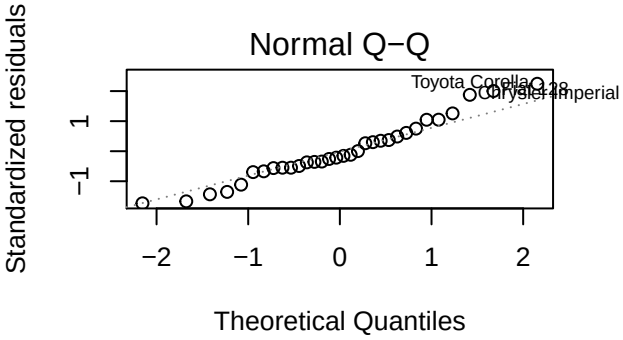
Figures

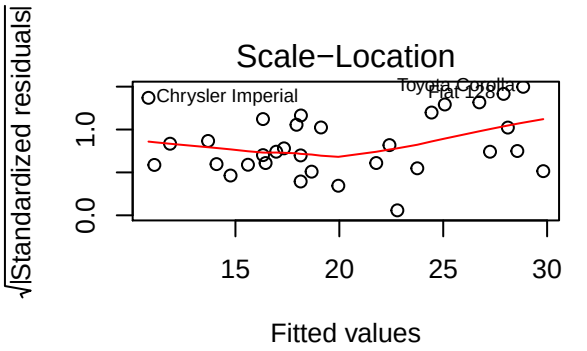
I

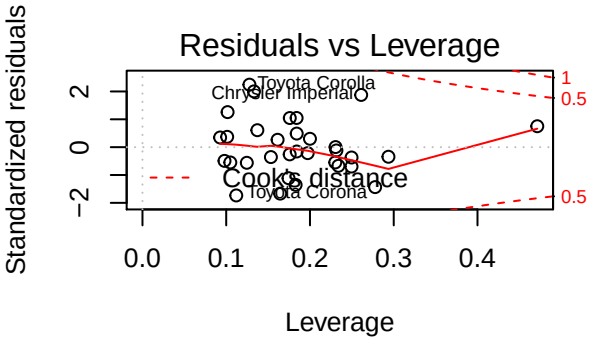

Milage by Transmission Type
Milage
10
15
20
25
30
0
1
Transmission Type (0 = Automatic, 1 = Manual)

II


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
15
20
25
30
−4
0
4


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Chrysler Imperial
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
15
20
25
30
0.0
1.0


Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5
0.0
0.1
0.2
0.3
0.4
−2
0
2