Motor MPG Analysis

Pratik Patil

04/11/2020

Report Summary The aim of this reoprt is to analyze the relation bewteen the transmission type of the vehicle(automatic or manual)and the distance(in miles) it covers per gallon(mpg). The final result aims at determining which transmission type gives more miles per gallon rating. The `mtcars` dataset is used to make the analysis.

Load Data Load the dataset and convert categorical variables to factors.

```
library(ggplot2)
data(mtcars)
head(mtcars, n=3)
dim(mtcars)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
attach(mtcars)
```

Exploratory Analysis **See Appendix Figure I**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 transmission type and MPG

```
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 Analysis Fit the full model of the data

```r
fullModelFit <- lm(mpg ~ ., data = mtcars)
summary(fullModelFit)  # results hidden
summary(fullModelFit)$coeff  # results hidden
```

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

Backward selection to determine which variables are most statistically significant

```r
stepFit <- step(fullModelFit)
summary(stepFit) # results hidden
summary(stepFit)$coeff # results hidden
```

The new model has 4 variables (cylinders, horsepower, weight, transmission). The R-squared value of 0.8659 confirms 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. Further increasing the cylinders to 8 with decrease the MPG by 2.16. Increasing the horsepower is decreases MPG 3.21 for every 100 horsepower. Weight decreases the MPG by 2.5 for each 1000 lbs increase. A Manual transmission improves the MPG by 1.81.

Residuals & Diagnostics Residual Plot **See Appendix Figure II**

The plots conclude:

A: The randomness supports the assumption of independence
B: The distribution of residuals is normal
C: Confirms the constant variance assumption
D: Since all points are under 0.05, there are no outliers

```r
sum((abs(dfbetas(stepFit)))>1)
```

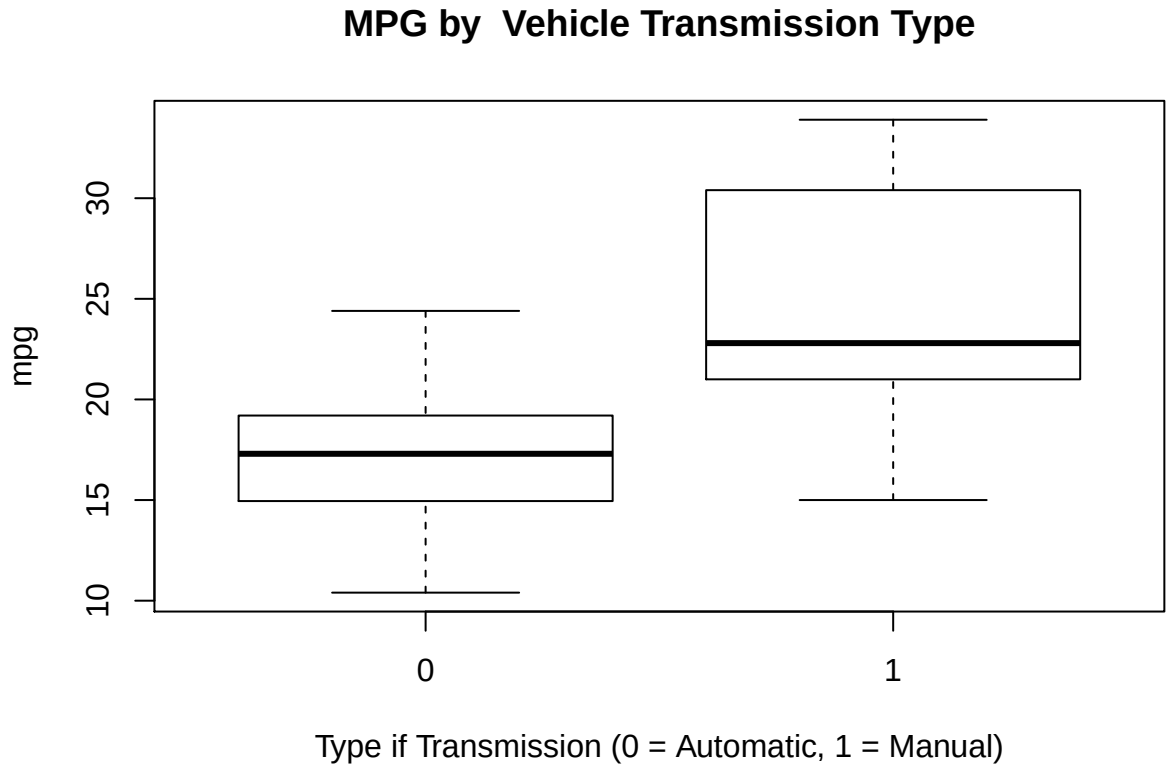
```
## [1] 0
```

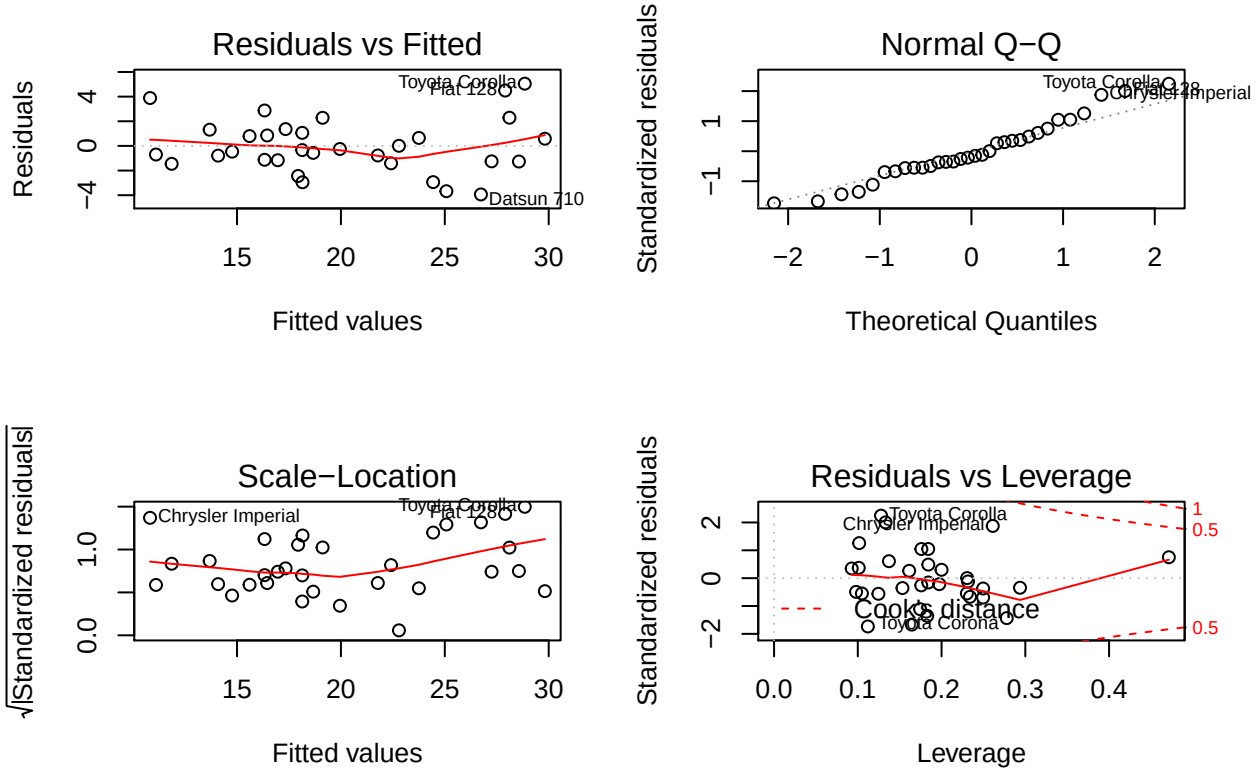
Results The analysis performed shows that the vehicles with manual transmission have a higher MPG rating than vehicles with automatic transmission by 7.245. When fitted with multiple linear regressions, the results went on to show that the transmission type had less to do with the earlier results, that is, the manual transmission itself contributed less to MPG, still giving it an edge over automatic transmission, but by 1.81 MPG. Other variables like it's enigne capacity(horsepower), it's weight and number of cylinders had a greater effect on the MPG of vehicles.

Conclusion There is a difference in MPG based on transmission type.

Appendix Figures

I Exploratory Box graph that compares Automatic and Manual transmission MPG.




Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Chrysler Imperial
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Toyota Corolla
Chrysler Imperial
Toyota Corona
1
0.5
0.5

II