Regression Models Project - Motor Trend MPG Data Analysis

Banu Priya

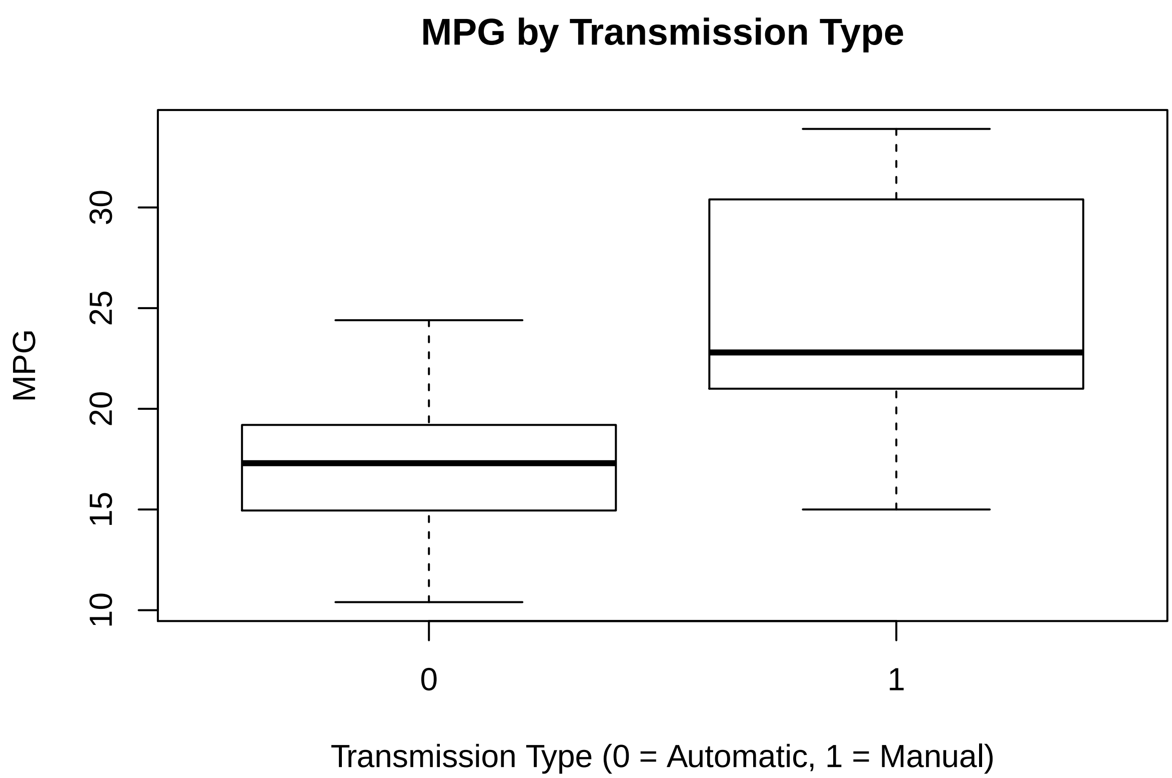
Aug 18, 2019

##– Executive Summary –> This report showcases the analysis of the relationship between transmission type (manual or automatic) and miles per gallon (MPG). The report is set out to determine which transmission type produces a higher MPG. The `mtcars` dataset is used for this analysis. A t-test between automatic and manual transmission vehicles shows that the manual transmission vehicles have a 7.245 greater MPG than the automatic transmission vehicles. On fitting the multiple linear regressions, the analysis shows that the manual transmission contributed less significant to MPG & also an improvement of 1.81 MPG.

##– 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 –> Exploratory Box graph that compares Automatic and Manual transmission MPG. 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

```
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.

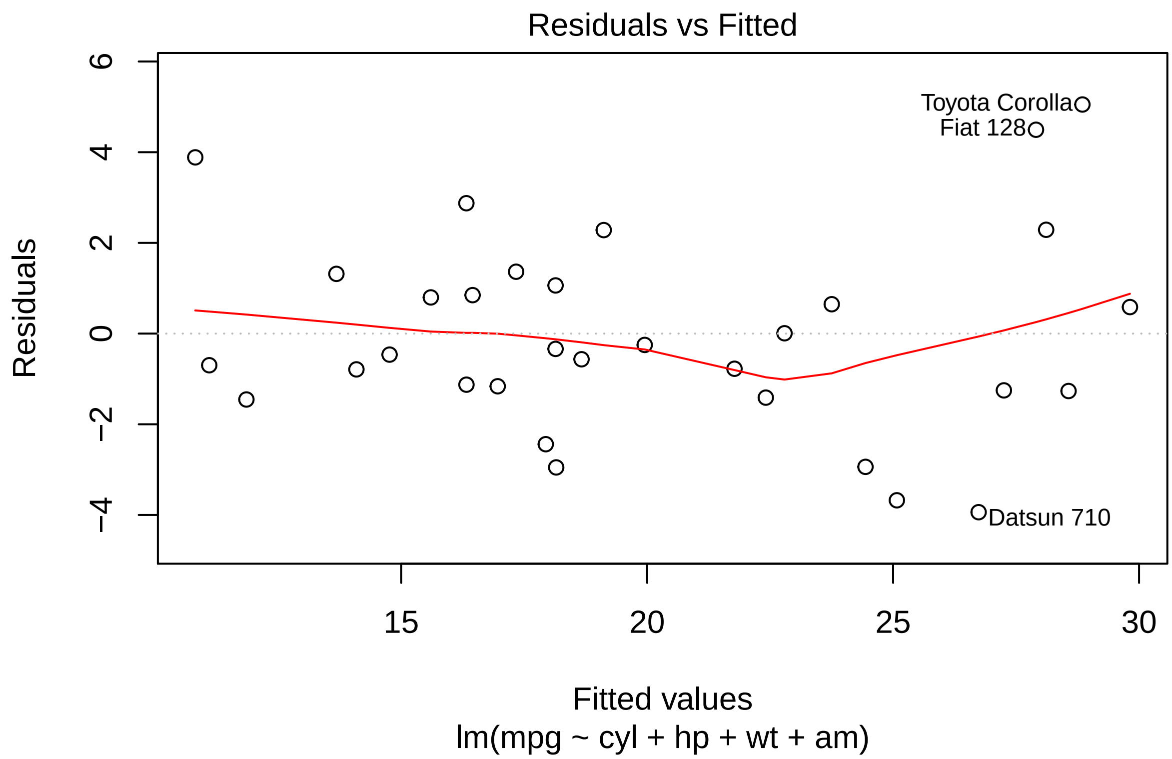
Backward selection to determine which variables are most statistically significant

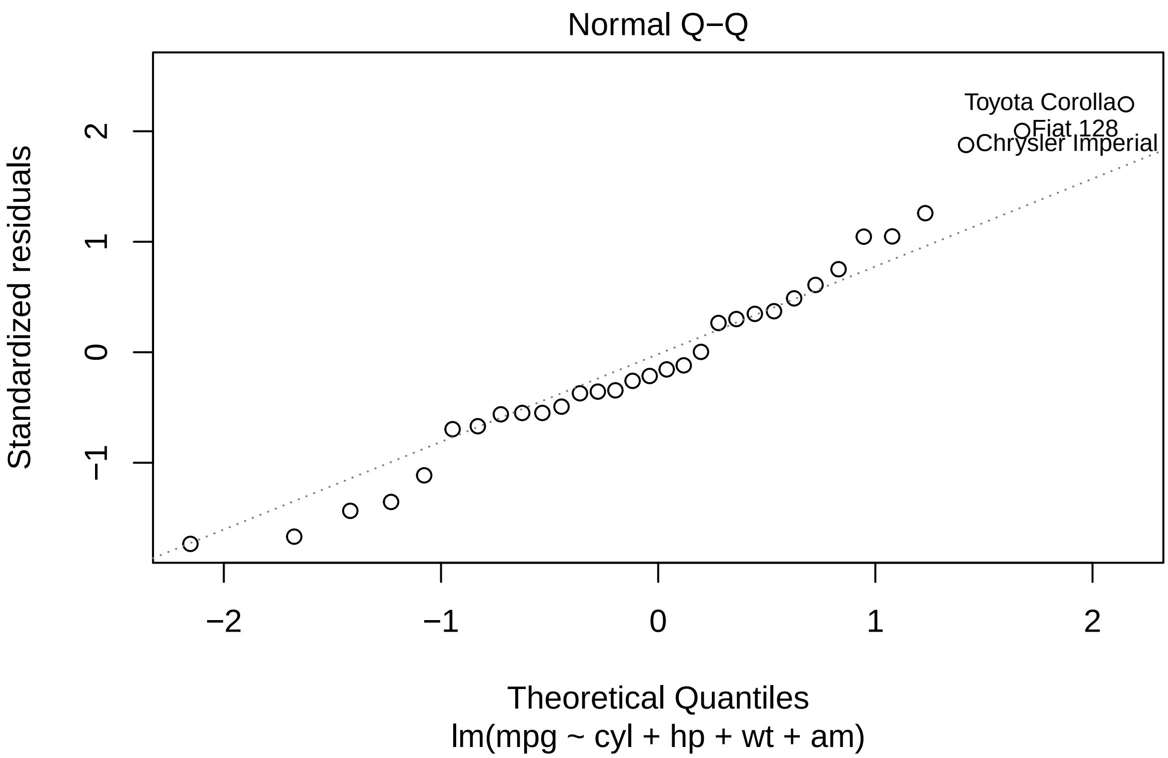
```r
stepFit <- step(fullModelFit)
summary(stepFit) # results hidden
summary(stepFit)$coeff # results hidden
```

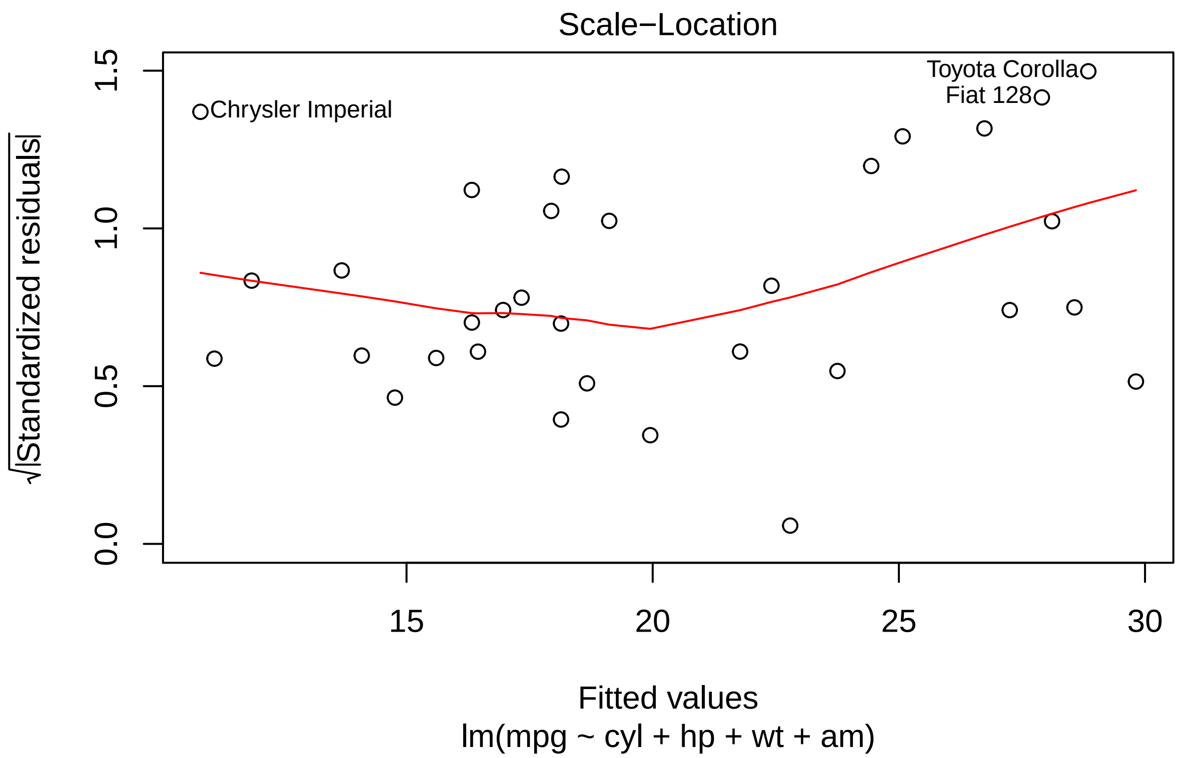
The new model includes 4 variables (cylinders, horsepower, weight, transmission). The R-squared value of 0.8659 confirms that this model explains about 87% of the variance in MPG. The p-values also are statistically significant because they have a p-value less than 0.05. The coefficients conclude that 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.

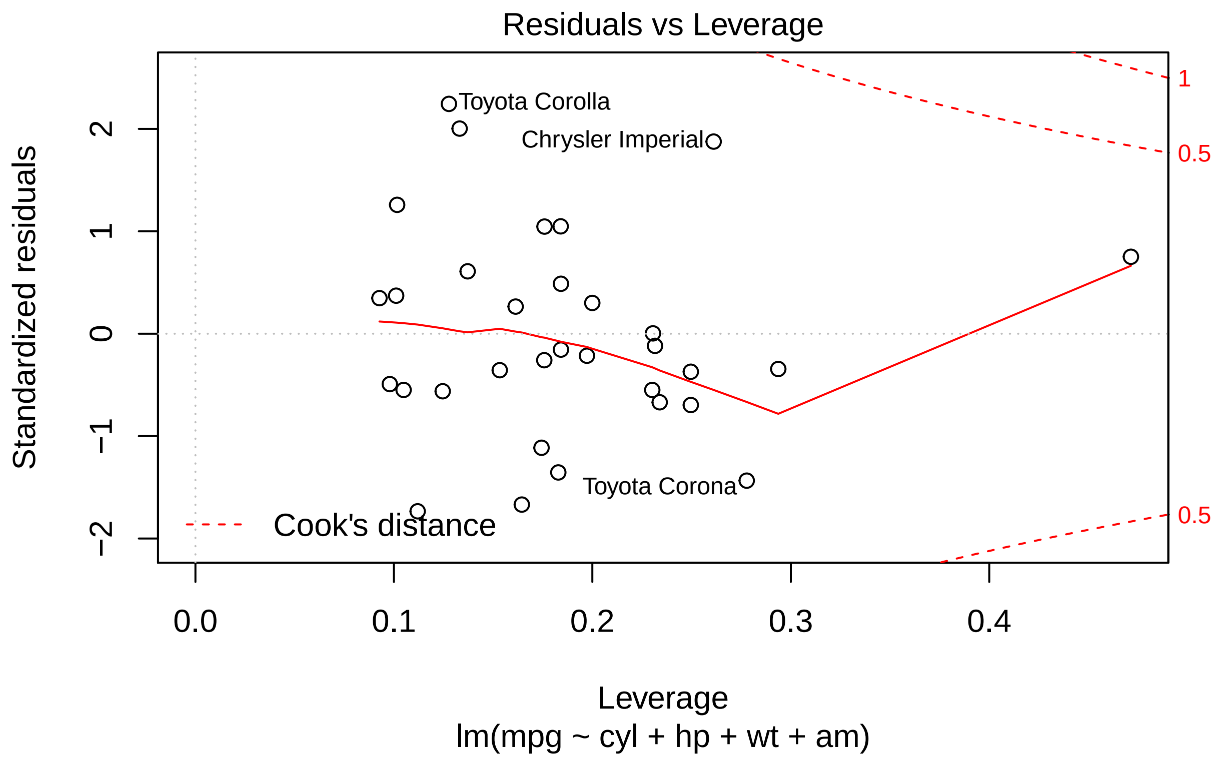
##– Residuals & Diagnostics –> Residual Plot

```
## Warning in par(mfrcol = c(2, 2)): "mfrcol" is not a graphical parameter
```




Normal Q−Q
Toyota Corolla
Fiat 128
Chrysler Imperial
Standardized residuals
−1
0
1
2
−2
−1
0
1
2
Theoretical Quantiles
lm(mpg ~ cyl + hp + wt + am)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
Chrysler Imperial
Toyota Corolla
Fiat 128
15
20
25
30
Fitted values
lm(mpg ~ cyl + hp + wt + am)



The plots conclude:

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

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

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

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