Regression Models Course Project

krrish16

16 September 2018

Executive Summary

Motor Trend is an automobile magazine in the United States for producing accurate automobile data to the public. The magazine known as Motor Trend (MT) covers the automobile industry with an interest in seeing the the relationship of MPG (miles per gallan) and a number of variables. The following questions need to be considered:

1.“Is an automatic or manual transmission better for MPG”

2.“Quantify the MPG difference between automatic and manual transmissions”

Loading data

```
library(ggplot2)
data("mtcars")
head(mtcars)
str(mtcars)
```

Processing data

```
#missing values check
colSums(is.na(mtcars))
```

```
##  mpg  cyl disp   hp drat   wt qsec   vs   am gear carb
##    0    0    0    0    0    0    0    0    0    0    0
```

```
#Transforming
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am   <- factor(mtcars$am,labels=c("Auto","Man"))
```

To observe relationships between the variables and, in particular, between mpg and am - the scatterplots produced by plotting each variable against all others, as well as the specific distribution of mpg values with am.

From the Exploratory Box graph which shows that there is a significant increase in MPG for vehicles with a manual transmission than with a automatic transmission.

Exploratory Data Anlysis

Performing T-test for transmisson type vs MPG:

We make the null hypothesis as the MPG of the automatic and manual transmissions are from the same population (assuming the MPG has a normal distribution). We use the two sample T-test to show it.

```
res <- t.test(mpg~am,data = mtcars)
res
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Auto  mean in group Man
##           17.14737           24.39231
```

From the p–value of the t-test results

1.We can reject the null hypothesis that the difference between transmission types is Null

2.The estimate values between the 2 transmissions is in favor of manualtranmission type.

Regression Analysis

```
fitting <- lm(mpg ~ ., data = mtcars)
summary(fitting) #results = "hide"
```

Since coefficients have a p-value less than 0.05 we cannot conclude which variables are more statistically significant.

```
fitting2 <- step(fitting) #results = "hide"
summary(fitting2) #results = "hide"
```

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.

Residual Analysis

The plots are a graphical examination of the residual analysis where we can infer

1.The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption.

2.The Normal Q-Q plot indicates that the residuals are normally distributed because the points lie closely to the line.

3.The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed.

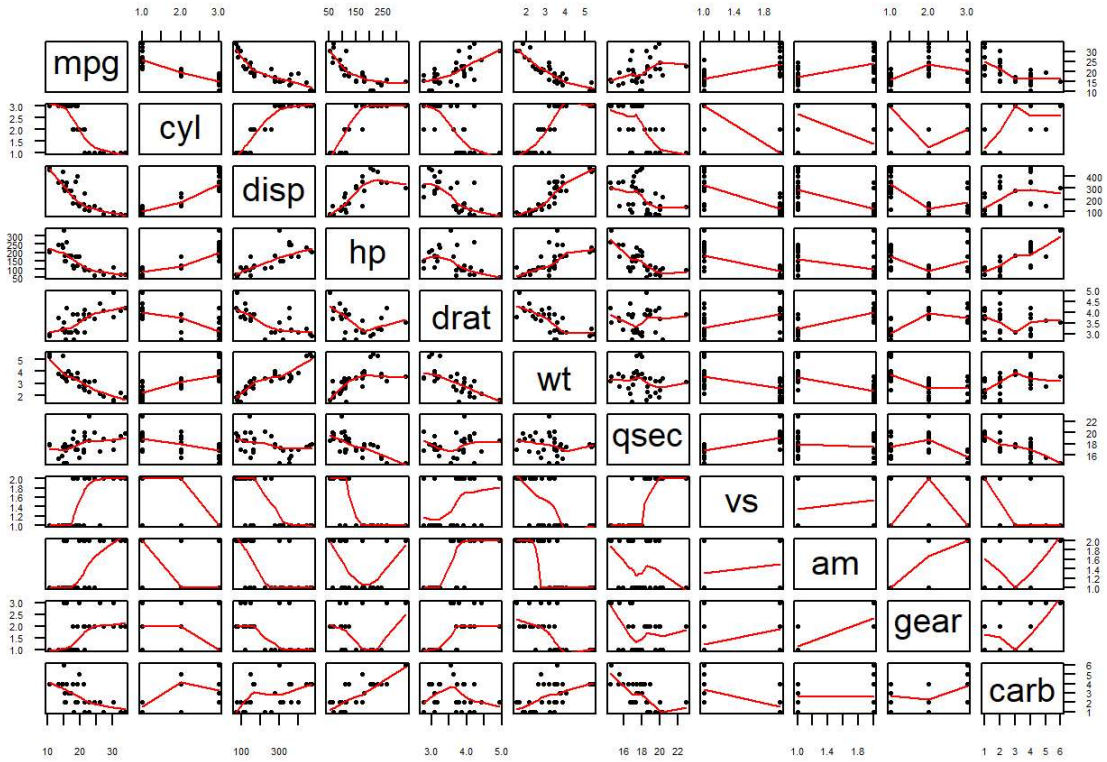
4.The Residuals vs. Leverage argues that no outliers are present, as all values fall well within the 0.5 bands.

Appendix

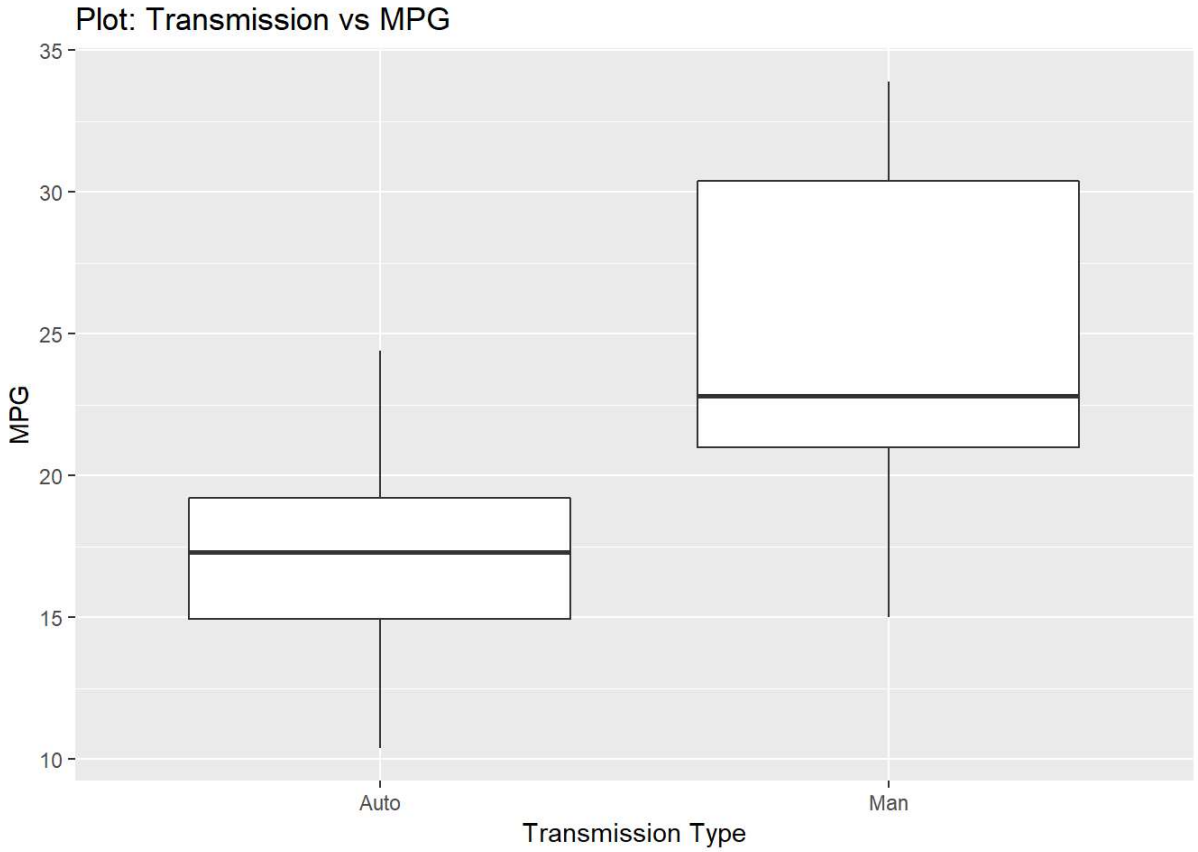
```
pairs(mtcars,panel=panel.smooth, pch=16, cex=0.5, gap=0.5, lwd=1, las=1, cex.axis=0.5)

library(ggplot2)
```

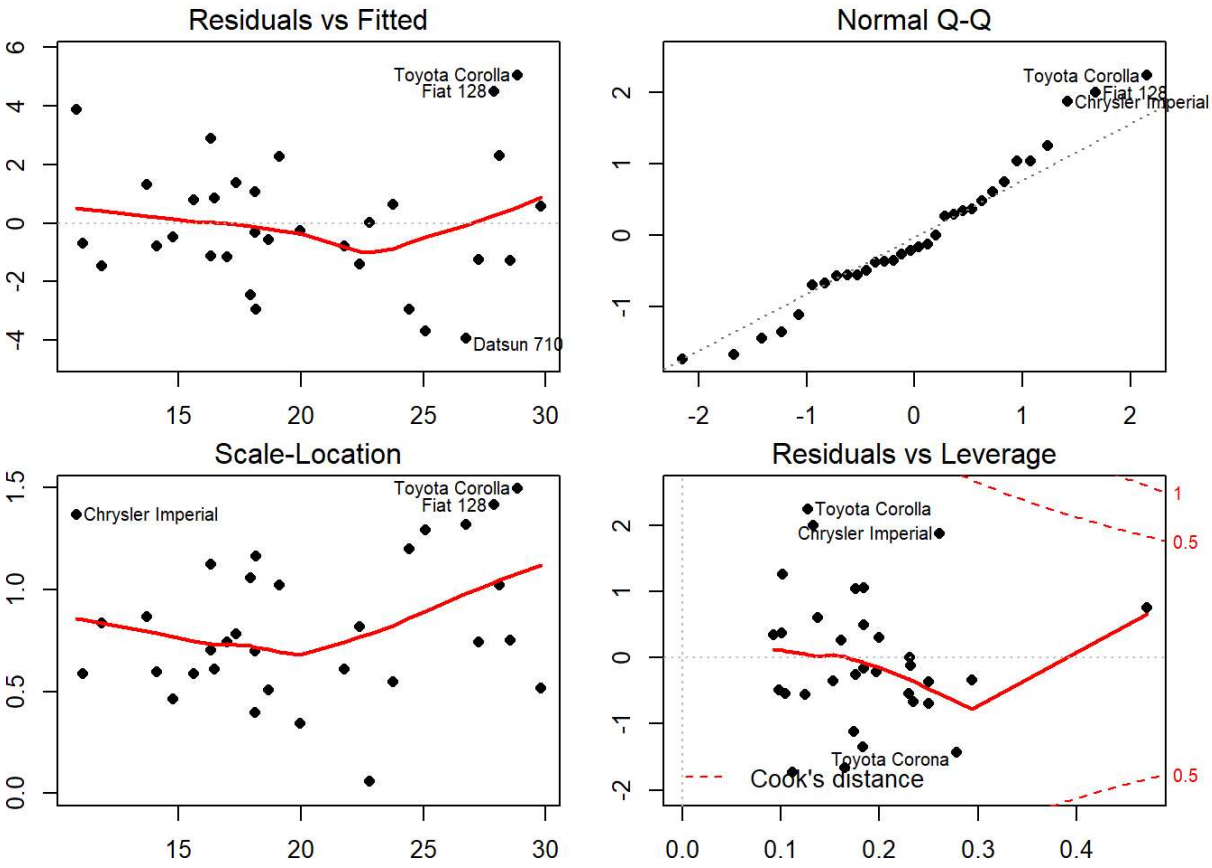
```
## Warning: package 'ggplot2' was built under R version 3.5.1
```




```
ggplot(data = mtcars, aes(am,mpg)) + geom_boxplot() + labs(x= "Transmission Type", y = "MPG", title = "P
lot: Transmission vs MPG")
```




```
# Residual Analysis Plots
par(mfrow=c(2,2), mai=.3*c(1,1,1,1))
plot(fitting2,pch=16,lty=1,lwd=2)
```


Residuals vs Fitted
Toyota Corolla
Fiat 128
Datsun 710
Normal Q-Q
Toyota Corolla
Fiat 128
Chrysler Imperial
Scale-Location
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5