MPG difference between automatic and manual cars

Basudha Misra

March 10, 2017

Executive Summary:

In this article we are interested to do an analysis to answer the following two questions:

- Is an automatic or manual transmission better for MPG(miles per gallon)?
- Quantify the MPG difference between automatic and manual transmissions.

We perform some exploratory analysis first to get a rough idea about the transmission effect on mpg. Based on this initial study we explore various regression models(with single and multivariable) and choose the best fitted model.

Basic exploratory analysis:

```
data(mtcars) # loading the data
#tranforming transmission type in terms of names
mtcars$am <- factor(mtcars$am,labels = c("Automatic","Manual"))
sum <- aggregate(mpg ~ am, data = mtcars, FUN = summary) # summary of the data
sum
```

```
##          am mpg.Min. mpg.1st Qu. mpg.Median mpg.Mean mpg.3rd Qu. mpg.Max.
## 1 Automatic    10.40       14.95      17.30    17.15       19.20    24.40
## 2    Manual    15.00       21.00      22.80    24.39       30.40    33.90
```

It can be clearly seen from both the summary and Fig1(Appendix), that MPG is much higher for automatic transmission cars than manual transmission. Next our aim is to quantify this difference with the help of statistical analysis.

Inference analysis:

MPG difference between automatic and manual type of transmission is 7.24 as shown below:

```
sum$mpg[2,4] - sum$mpg[1,4]
```

```
## Mean
## 7.24
```

We perform a t-test analysis to check the significance of this difference.

```
Tautomatic <- mtcars[mtcars$am == "Automatic",]
Tmanual <- mtcars[mtcars$am == "Manual",]
t.test(Tautomatic$mpg, Tmanual$mpg)$p.value
```

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

As p-value is 0.14% which is significantly less than 5%, we can say from this statistical inference analysis that the mpg difference between automatic and manual transmission is significant. We perform regression analysis in next section to properly quantify this difference.

Regression Analysis:

First we perform a basic regression.

```
fitbasic <- lm(mpg ~ am, data = mtcars)
summary(fitbasic)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fitbasic)$r.squared
```

```
## [1] 0.3597989
```

This model shows that the average MPG due to manual transmission is 7.24 higher than the average MPG 17.15 of automatic transmission. R^2 value suggests that this model can explain only 36% of the variance which indicates that we should explore multivariate regression models. To build the best multivariate regression model, we first explore the correlation between mpg and various other variables of mtcars data in Fig2(higher correlation with MPG) and Fig3(lower correlation with MPG) which can be seen in Appendix. Based on this analysis we figure out the best fitted multivariate model.

```
fitbest <- lm(mpg ~ am + wt + cyl + disp + hp, data = mtcars)
anova(fitbasic,fitbest) # coparison with basic model
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + cyl + disp + hp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 163.12  4    557.78 22.226 4.507e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This anova analysis shows that the multivariate model is much better than our basic model.

```
summary(fitbest)$coef[1:2,]
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 38.202799   3.669096 10.412045 9.084987e-11
## amManual     1.556492   1.440536  1.080495 2.898430e-01
```

```
summary(fitbest)$r.squared
```

```
## [1] 0.8551394
```

Above result shows that MPG of manual transmission is 1.56 larger than automatic transmission MPG and this best fitted multivariate model explains 85.51% of the variance. For further confirmation we do a double check of the normality test of the residuals in Fig4(Appendix) and this explains that the residuals are normally distributed. It further supports our best fitted model.

Conclusion(Quantification):

Model	difference between automatic and manual(MPG)	Variance Explained(R^2)
Basic(fitbasic)	7.24	35.98%
Bestfitted(fitbest)	1.56	85.51%

Appendix:

Details of the mtcars data can be seen from here.

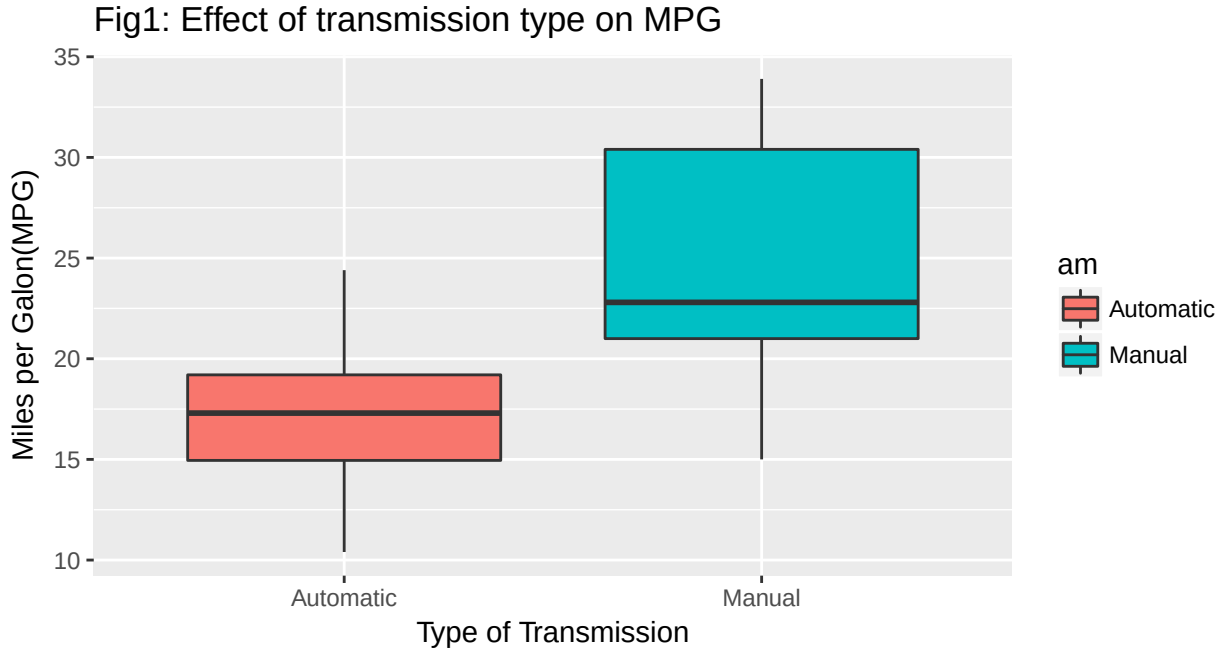
In this section we present various plots and their codes which are used to produce this report.

Exploratory analysis plot:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```
gexp <- ggplot(mtcars, aes(x = am, y = mpg, fill = am)) +
          geom_boxplot() +
            labs(x = "Type of Transmission", y = "Miles per Galon(MPG)",
                   title = "Fig1: Effect of transmission type on MPG")
 print(gexp)
```



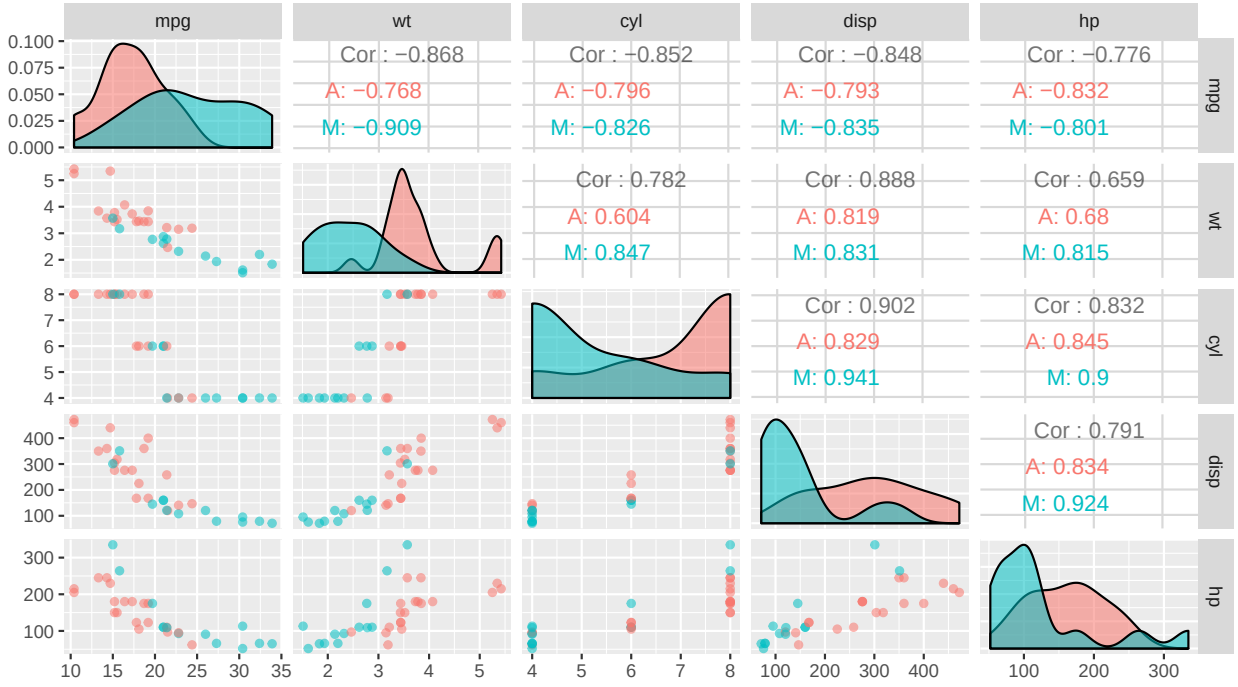
Pairwise plot:

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.3.3
```

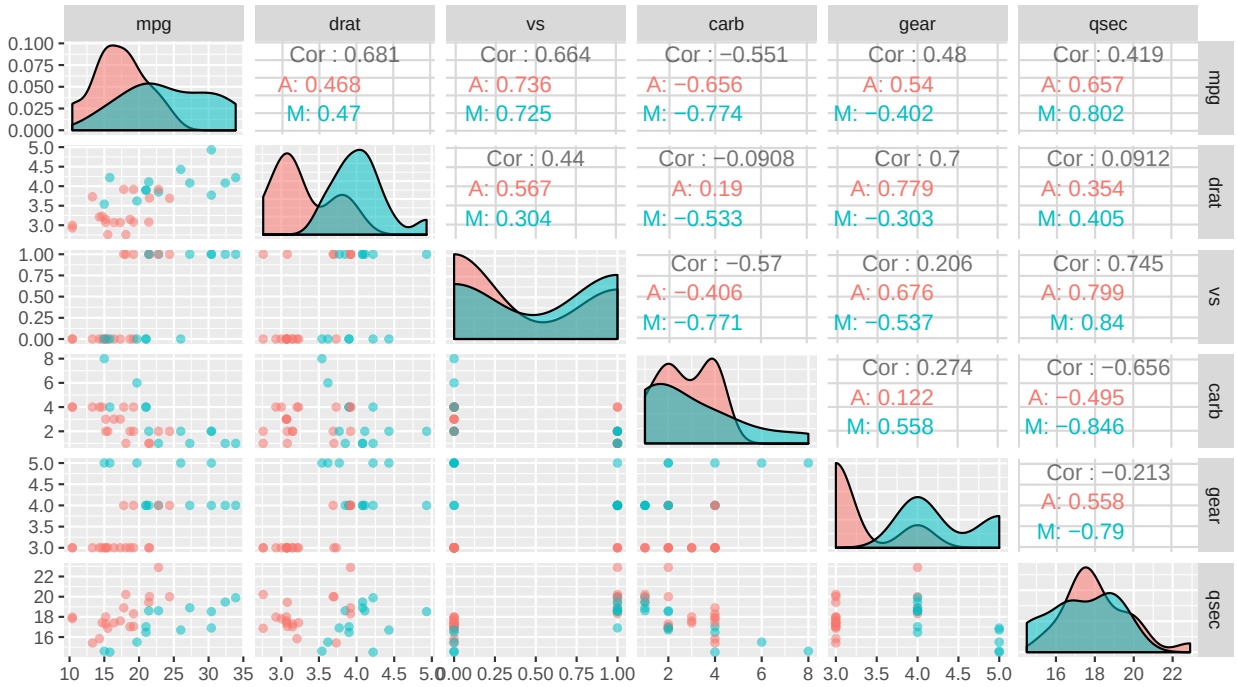
```
mtcars$am = factor(mtcars$am, labels = c("A","M"))
ggpairs(mtcars, mapping=aes(colour=am, alpha = 0.7),
        columns = c("mpg","wt","cyl","disp","hp")) +
          labs(title = "Fig2: Higher correlated variables with MPG")
```

Fig2: Higher correlated variables with MPG




```
ggpairs(mtcars, mapping=aes(colour=am, alpha = 0.7),
        columns = c("mpg","drat","vs","carb","gear","qsec"))+
          labs(title = "Fig3: Lower correlated variables with MPG")
```

Fig3: Lower correlated variables with MPG



Residuals diagnostic plots:

```
par(mfrow=c(2,2),oma = c(0, 0, 2, 0))
plot(fitbest)
```

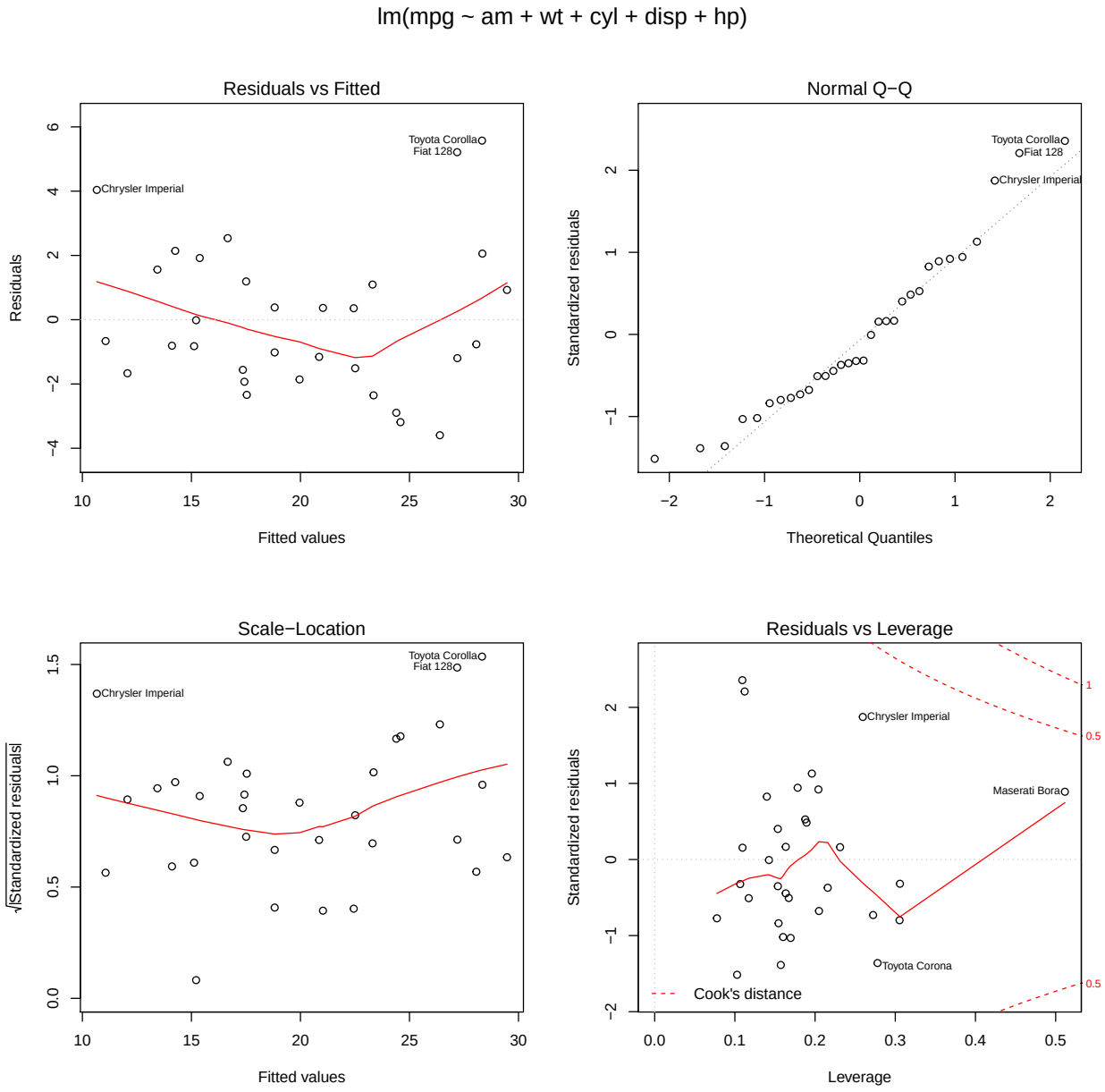


Fig4: Diagnostic Plots