# Report for Motor Trend

*Yilun Zhang*

*Tuesday, November 04, 2014*

**Objectives** This is the report for Motor Trend, a magazine about the automobile industry. This report aims to answer the following two questions:

1. Is an automatic or manual transmission better for MPG
2. Quantification of the MPG difference between automatic and manual transmissions

## Load and view data

```
data <- mtcars
head(data)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

## Modification of data

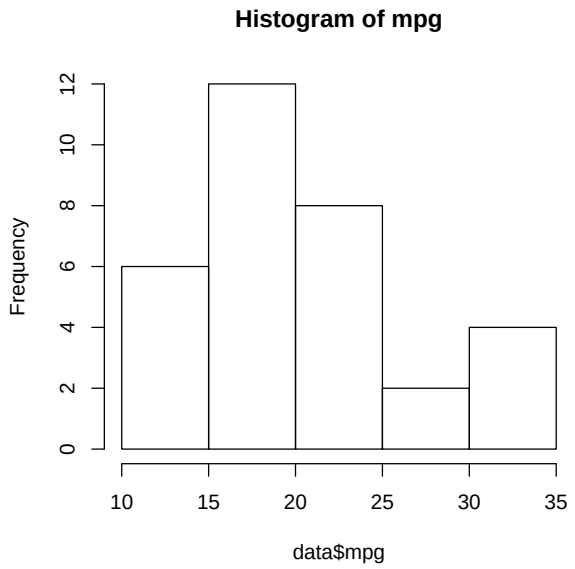
We noticed that the input data has all columns being numberic, the following code will modify some of the data types to factors in order to process further analysis.

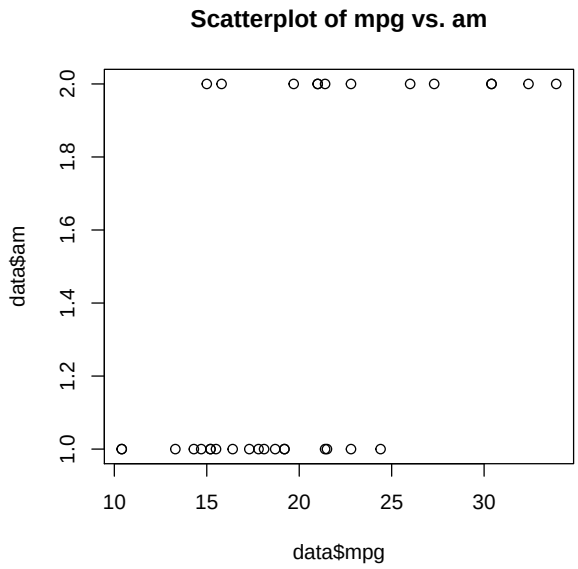
```
data$cyl <- factor(data$cyl)
data$vs <- factor(data$vs)
data$am <- factor(data$am)
data$gear <- factor(data$gear)
data$carb <- factor(data$carb)
```

## Exploratory data analysis

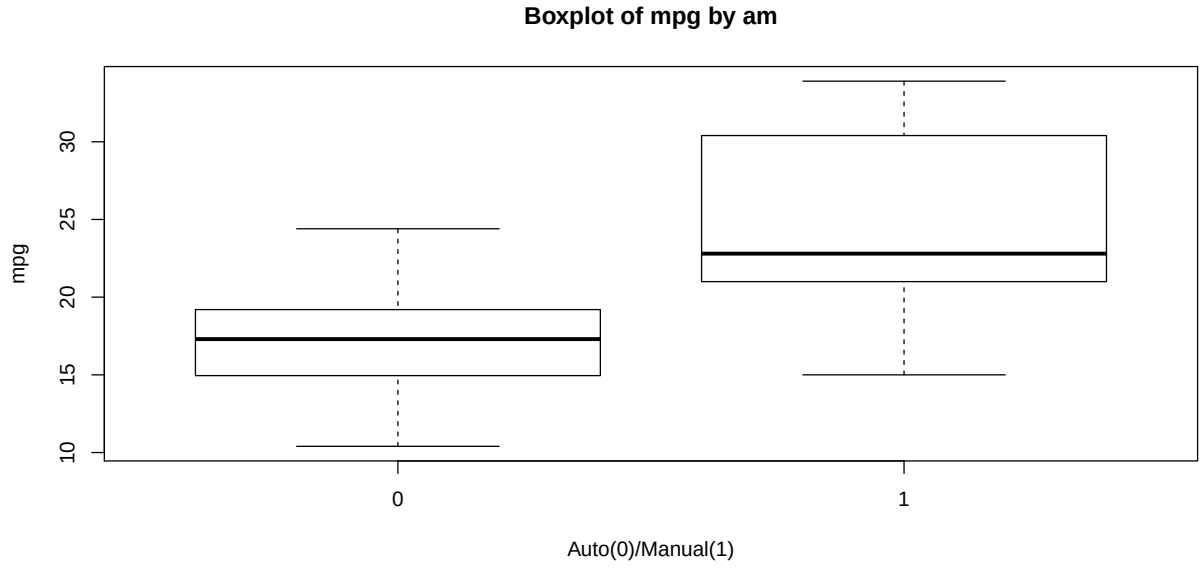
First, we will conduct some exploratory analysis on our data to get familiarize with it. Mainly, we are looking at mpg and am.

```
par(mfrow=c(1,2))
hist(data$mpg,main="Histogram of mpg")
plot(data$mpg,data$am,main="Scatterplot of mpg vs. am")
```




```
par(mfrow=c(1,1))
boxplot(data$mpg~data$am,main="Boxplot of mpg by am",xlab="Auto(0)/Manual(1)",ylab="mpg")
```




## Question 1

By looking at the figures above, we noticed that, from the boxplot, manual transmission has a higher median mpg value than automatic transmission. We conclude at this point that manual transmission has a better mpg than automatic transmission

## Question 2

Let's formally analyse if there is a difference between the two types of transmissions on mpg.

```
mean(data[data$am == 0,]$mpg)
```

### Mean Comparison

```
## [1] 17.14737
```

```
mean(data[data$am == 1,]$mpg)
```

```
## [1] 24.39231
```

Manual transmission has a mean value of 24.39231 but automatic transmission only has a mean of 17.14737.

```
t.test(data[data$am == 0,]$mpg,data[data$am == 1,]$mpg,alternative="less")
```

### T Test

```
##
##  Welch Two Sample t-test
##
## data:  data[data$am == 0, ]$mpg and data[data$am == 1, ]$mpg
## t = -3.7671, df = 18.332, p-value = 0.0006868
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -3.913256
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

The one sided t-test retuen a p-value of 0.001374, which rejected the null hypothesis of no difference and favors that manual transmission has a better mpg than automatic transmission.

```
fit1 <- lm(mpg~am,data=mtcars)
summary(fit1)$coef
```

### Regression modelling

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

```
fit2 <- lm(mpg~am-1,data=mtcars)
summary(fit2)$coef
```

```
##    Estimate Std. Error  t value     Pr(>|t|)
## am 24.39231   3.956183 6.165616 7.666189e-07
```

We have fitted two models:

- fit1 has an intercept and the slope
- fit2 has only the slope (goes throuth the origin)

```
summary(fit1)$sigma
```

```
## [1] 4.902029
```

```
summary(fit2)$sigma
```

```
## [1] 14.26422
```

By looking at the coefficient summary of all three models, auto/manual is significantly important in both models (all have p-value <0.05).

By looking at the residual standard errors of the two SLRs (Simple Linear Regression), fit1 has a smaller value.

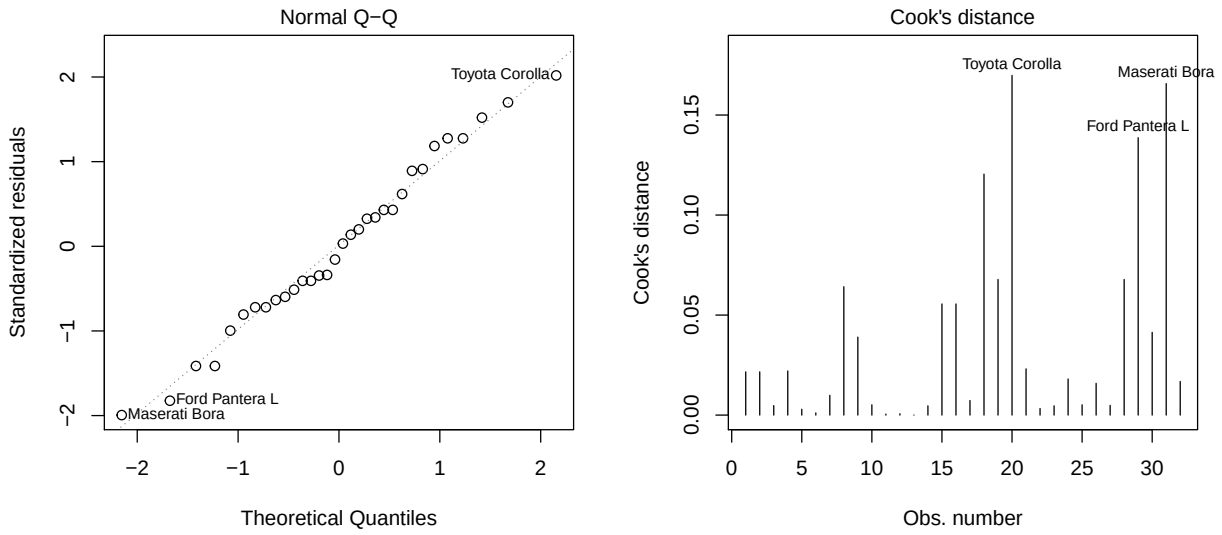
```
anova(fit1,fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am - 1
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30  720.9
## 2     31 6307.5 -1   -5586.6 232.49 1.134e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova F test returns an f-value of less than 0.05, which indicates that two models are significantly different, so reducing the intercept causes an sigficant change in the model. Therefore, fit1 is the better model.

Finally, let's check the residual of fit1.

```
par(mfrow=c(1,2))
plot(fit1,which=2)
plot(fit1,which=4)
```




```
shapiro.test(fit1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit1$residuals
## W = 0.9821, p-value = 0.8573
```

From thel left QQ plot, we can see that the residuals are approximately normal, the Shapiro-Wilk test also confirms that the residuals are normal (p-value of larger than 0.05).

The cook's distance plot shows that there are no extreme values in residual, so the residual looks quite good.

## Summary

In conclusion, by comparing the mean mpg of auto/manual, fitting regression models on them, we conclude that manual transmission has better mpg than auto transmission.