Regression Models Project Report

Xiaoyi Liu

2015-07-25

Executive Summary

In this project, we are going to explorer the relationship between a set of variables and miles per gallon (MPG) (outcome) in the "mtcars" dataset. We are particularly interested in the following two questions:

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

We are going to fit different linear models to discover the influences of automatic or manual transmission towards mpg conditioned on different features.

Data Processing

```
data(mtcars)
dim(mtcars)
```

```
## [1] 32 11
```

```
head(mtcars, 3)
```

```
##                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
```

We can see the "mtcars" dataset consists of 32 samples, each with 11 features.

Data Analysis

In this section, we are going to explorer the answers for the two questions listed in the abstract section.

First, we would like to find out that whether automatic or manual transmission is better for MPG. We present the boxplot of MPG with respect to automatic or manual in the appendix.

It is quite clear that automatic is much better than manual transmission. Let's verify this through more rigorous data analysis.

```
mean(subset(mtcars$mpg, mtcars$am == 1)) - mean(subset(mtcars$mpg, mtcars$am == 0))
```

```
## [1] 7.244939
```

We can see the automatic transmission has a much higher mean than the manual ones.

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

The t-test result shows a p-value of 0.001374, which is much smaller than 0.05. Apparently, we should accept the alternative hypothesis that there exists true difference in mean between automatic and manual transmissions. Therefore, we should conclude that automatic is better than manual in MPG. Next, we are going to quantify the MPG difference between them.

```
mtcars$am = as.factor(mtcars$am)
fit1 = lm(mpg ~ ., data = mtcars)
summary(fit1)
```

We fit a linear model for mpg based on all the features. It can be seen that all the p-values are significanlty larger than 0.05, which means the null hypothesis that the corresponding coefficient for each feature is zero should be accepted. We will select the features wt and qsec with the smallest p-values to analyze their interactions with am towards mpg.

```
fit2 = lm(mpg ~ wt + qsec + am, data = mtcars)
summary(fit2)
```

It can be seen that compared with qsec, the p-value for wt is much smaller than that of qsec, which suggests that the the null hypothesis that the coefficient for wt is zero is not supported. We should dig more into the effects of wt now.

```
fit3 = lm(mpg ~ am * wt, mtcars)
summary(fit3)
```

It can be seen that for a fixed weight, the manual transmission (represented by am1) can earn more mpg (14.8784 - 5.2984 $>$ 0) than automatic one, which is a little bit out of surprise.

The residual plot is shown in the appendix. We can observe that: (1) No consistent pattern appears in the figure of "Residuals vs Fitted"; (2) No outliers in the dataset are present, since all the values in the figure of "Residuals vs Leverage" fall into the 0.5 bands.

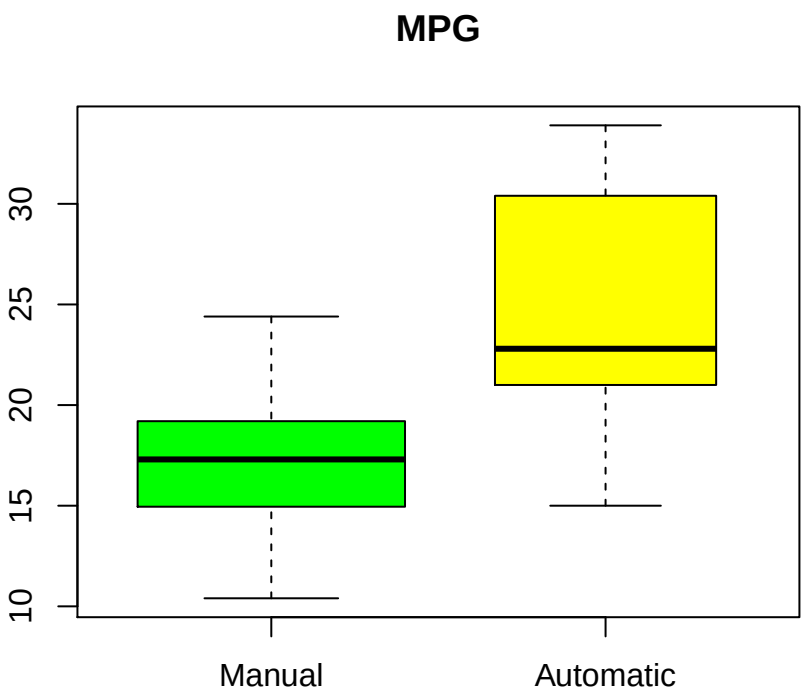
Conclusions

Through the data analysis towards the "mtcars" dataset, we have made the following conclusions:

1. Automatic transmission is better than the manual one for MPG.
2. With fixed weight, manual transmission will gain more mpg than the automatic transmission.

Appendix

```
boxplot(mpg ~ am, data = mtcars, varwidth = TRUE, col=c("green","yellow"), names = c("Manual","Automati
```




```
par(mfrow = c(2, 2))
plot(fit3)
```