Regression Models Course Project

Seongsu Kim

August 18, 2015

Executive Summary

- You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:
  - "Is an automatic or manual transmission better for MPG"
  - "Quantify the MPG difference between automatic and manual transmissions"

Mission

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

- The report must be:
  - Written as a PDF printout of a compiled (using knitr) R markdown document.
  - Brief. Roughly the equivalent of 2 pages or less for the main text.
  - Include a first paragraph executive summary.

Exploratory Data Analysis

1. To begin with, load the data.

```
data(mtcars)
```

2. Here are the details of each column

[, 1] mpg Miles/(US) gallon

[, 2] cyl Number of cylinders

[, 3] disp Displacement (cu.in.)

[, 4] hp Gross horsepower

[, 5] drat Rear axle ratio

[, 6] wt Weight (lb/1000)

[, 7] qsec 1/4 mile time

[, 8] vs V/S

[, 9] am Transmission (0 = automatic, 1 = manual)

[,10] gear Number of forward gears

[,11] carb Number of carburetors

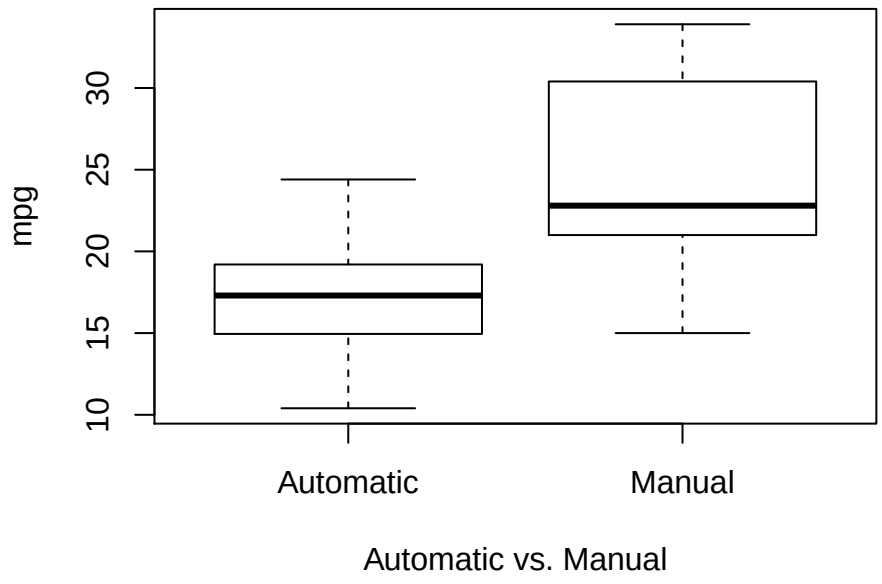
3. By using t-test, we can see the mpg diffence between a manual transmission and an auto transmission.

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

```
##
##  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 0 mean in group 1
##        17.14737        24.39231
```

According to the result shown above,a mean mpg in group 0 which is the group of an automatic transmission is 17.147. On the other hand, the mean mpg in group 1 which is the group of an manual transmission is 24.39231. However, we need some more analysis to confirm the result.

```
mtcars$am[mtcars$am=="0"]<-"Automatic"
mtcars$am[mtcars$am=="1"]<-"Manual"
boxplot(mpg ~ am, data = mtcars, ylab = "mpg", xlab = "Automatic vs. Manual")
```



4. Multiple Linear Regression & Residual Plot

```
summary(step(lm(formula = mpg ~ am + qsec+wt, data = mtcars),trace=0,steps=100000))
```

```
##
## Call:
## lm(formula = mpg ~ am + qsec + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## amManual      2.9358     1.4109   2.081 0.046716 *
## qsec          1.2259     0.2887   4.247 0.000216 ***
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The result above shows that Multiple R-squared is about 85%. It means mpg has 85% of variation. Also, the coefficient is 2.94. The coefficient in this result means the average of mpg of manual transmission cars is 2.94 higher than the automatic transmission does. However, this result doesn't show the relationship with the one we got on the second step. The one we got before was 7 which is more than twice bigger than 2.94. Also, we can't find heteroskedastic relationship from the residual plot. Therefore, I would conclude that it is not possible to determine the better type of transmission for mpg from the data given.

Appendix

Residual plot

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