Coursera Regression Course Project

EF

April 3, 2016

Executive Summary

The goal of this analysis is the determine whether there is a meaningful difference in the fuel efficiency for cars with automatic vs manual transmissions. Once this difference is established it will be formally quanitified. The dataset used is the "mtcars"" dataset built in to the R programming language.

Load Libraries and Data

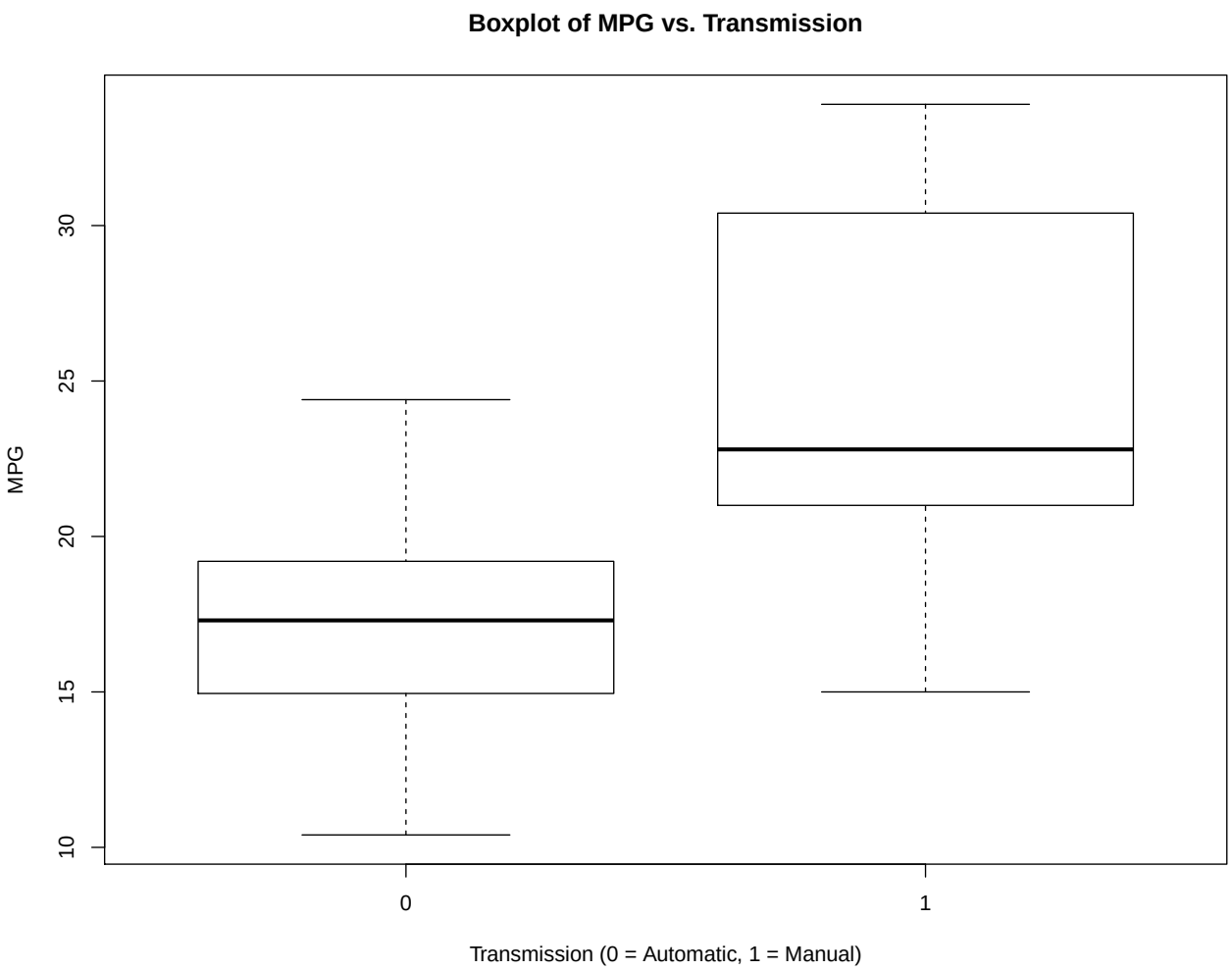
```
# load neccesary libraries
library(ggplot2)
library(datasets)
library(gridExtra)
library(GGally)

# load mtcars dataset
data(mtcars)
```

Exploratory Data Analysis

Since we are priarily concerned with the question of fuel economy as a function of automatic vs manual transmisisons cars, let's first create a boxplot to view the fuel economy values with the data split into two groups representing cars with automatic vs manual transmissions.

```
boxplot(mtcars$mpg ~ mtcars$am, xlab="Transmission (0 = Automatic, 1 = Manual)", ylab="MPG", main="Boxpl
```



From the figure above we can see a clear increase in fuel efficiency in the cars with manual transmissions. This plot answers the initial question of whether there is a difference in fuel economy for manual vs automatic transmissions; now we can proceed to quantifying this difference.

Regression Analysis

In the interest of parsimony, we will perform a simple linear regression model to investigate the relationship between fuel efficiency and transmission type.

```
fit <- lm(mpg ~ am, data=mtcars)
summary(fit)
```

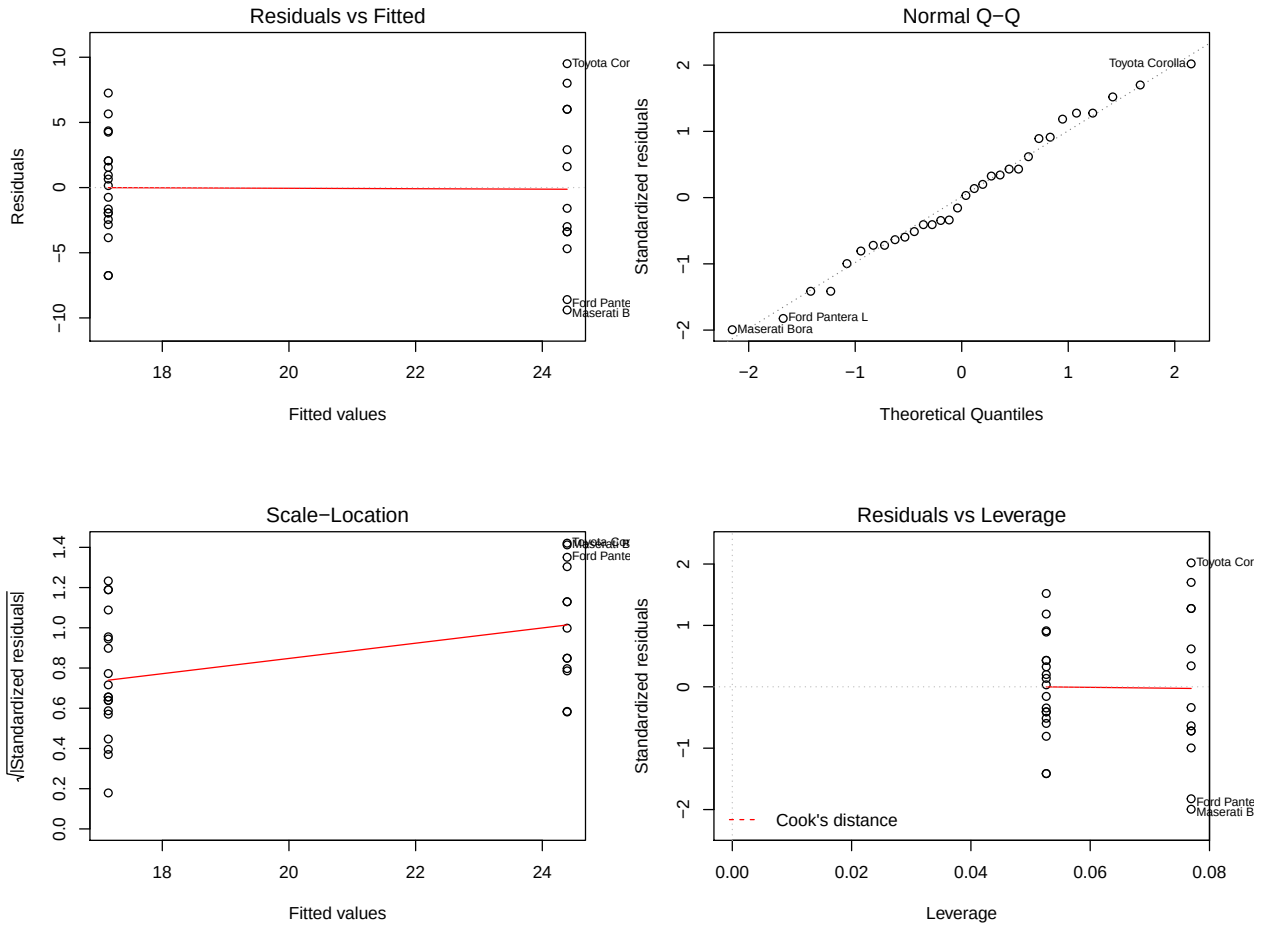
```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Our summary output shows that on average, a car has 17.147 mpg with an automatic transmission,with an increase of 7.245 mpg for manual transmissions. This model has the Residual standard error as 4.902 on 30 degrees of freedom. The Adjusted R-squared value is 0.3385, which means that the model can explain about 34% of the variance of the MPG variable. This is not particularly high.

Let's plot the residuals.

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



I know my model needs work and that there are other confounding parameters to work on going forward.