Regression Models Course Project- Car MPG Analysis

Daisy Wu

2017 年 12 月 25 日

Summary

Motor Trend is a automobile industry Magazine. We are interested the relationship between variables that affect miles per gallon MPG.

1. Are automatic or manual transmission better for MPG?
2. What are the MPG differences between automatic/manual transmissions?

Using a data set provided by Motor Trend Magazine do linear regression and hypothesis testing, to see if there is a significant MPG differences between automatic and manual transmission.

To quantify the MPG difference between automatic and manual transmission cars, a linear regression model was used to take into account the weight, transmission type and the acceleration. Based on these findings manual transmissions have better fuel economy of 2.94 MPG more than automatic transmissions.

Load Libraries

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Read the Data

```
data(mtcars)
str(mtcars)

## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
```

```
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Process the Data

convert "am" to a factor variable, "AT"= Automatic Transmission and "MT"= Manual Ttansmission.

```
mtcars$am<-as.factor(mtcars$am)
levels(mtcars$am)<-c("AT", "MT")
```

Exploratory Data Analysis

Get the Mean of Automatic and Manual Transmissions:

```
aggregate(mpg~am, data=mtcars, mean)
```

```
##   am      mpg
## 1 AT 17.14737
## 2 MT 24.39231
```

The mean MPG for manual transmissions is 7.245 which higher than automatic transmission cars. Is this significant?

Validate Significance:

```
aData <- mtcars[mtcars$am == "AT",]
mData <- mtcars[mtcars$am == "MT",]
t.test(aData$mpg, mData$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  aData$mpg and mData$mpg
## 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 of x mean of y
##  17.14737  24.39231
```

The p-value of the t-tst is 0.001374, with 95% confidence interval. There is a significant difference between the mean MPG for automatic verses manual transmissions.

Histogram of the mpg for Automatic and Manual Trasmissions.

```
ggplot(data = mtcars, aes(mpg)) + geom_histogram() + facet_grid(.~am) +
 labs(x = "Miles per Gallon", y = "Frequency", title = "MPG Histogram f
or automatic verses manual transmissions")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
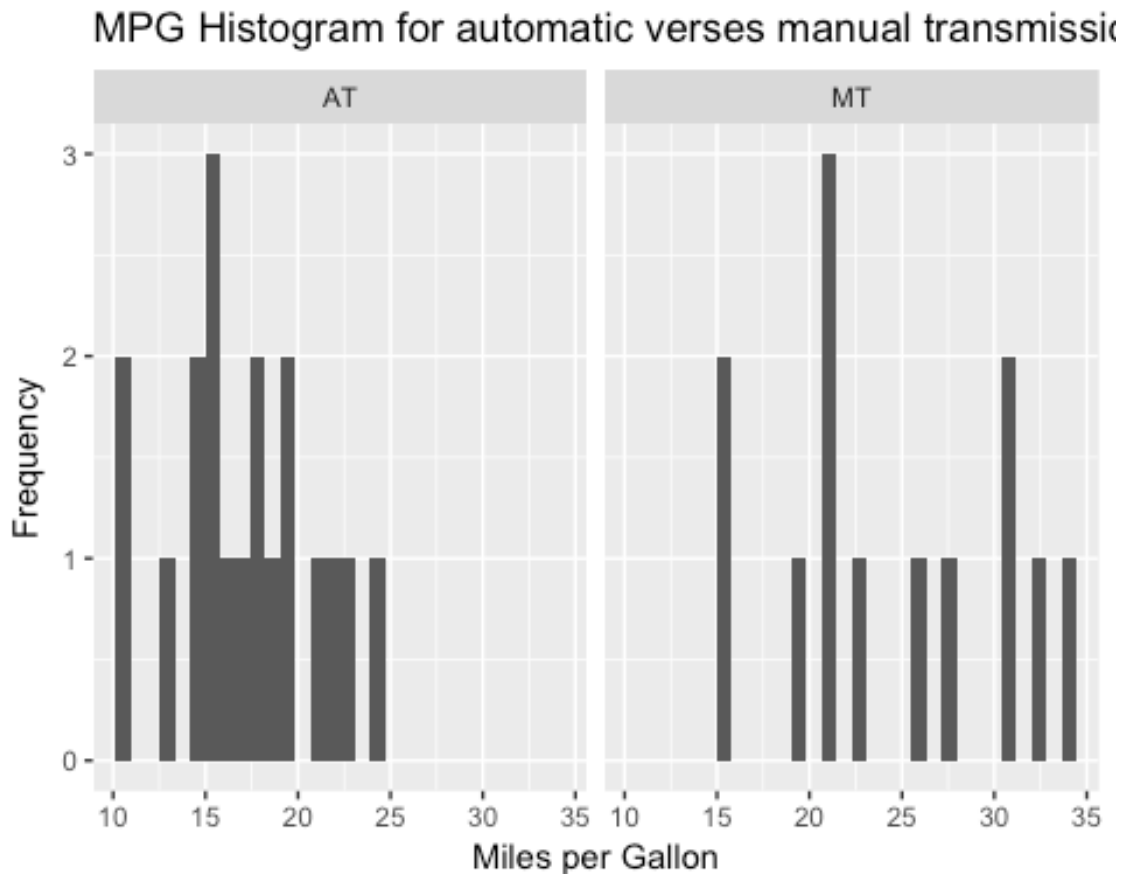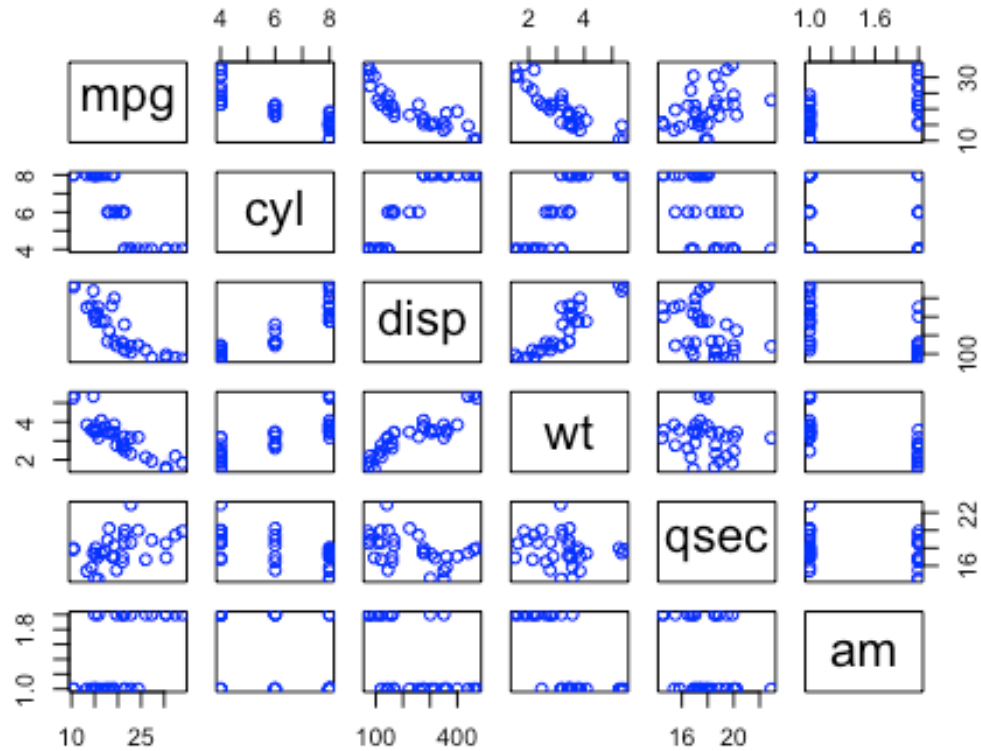

Boxplot mpg for Automatic and Manual Trasmissions

```
ggplot(data = mtcars, aes(am,mpg)) + geom_boxplot() + labs(x= "Transmis
sion", y = "MPG", title = "MPG: Automatic and Manual Trasmissions")
```

MPG: Automatic and Manual Trasmissions
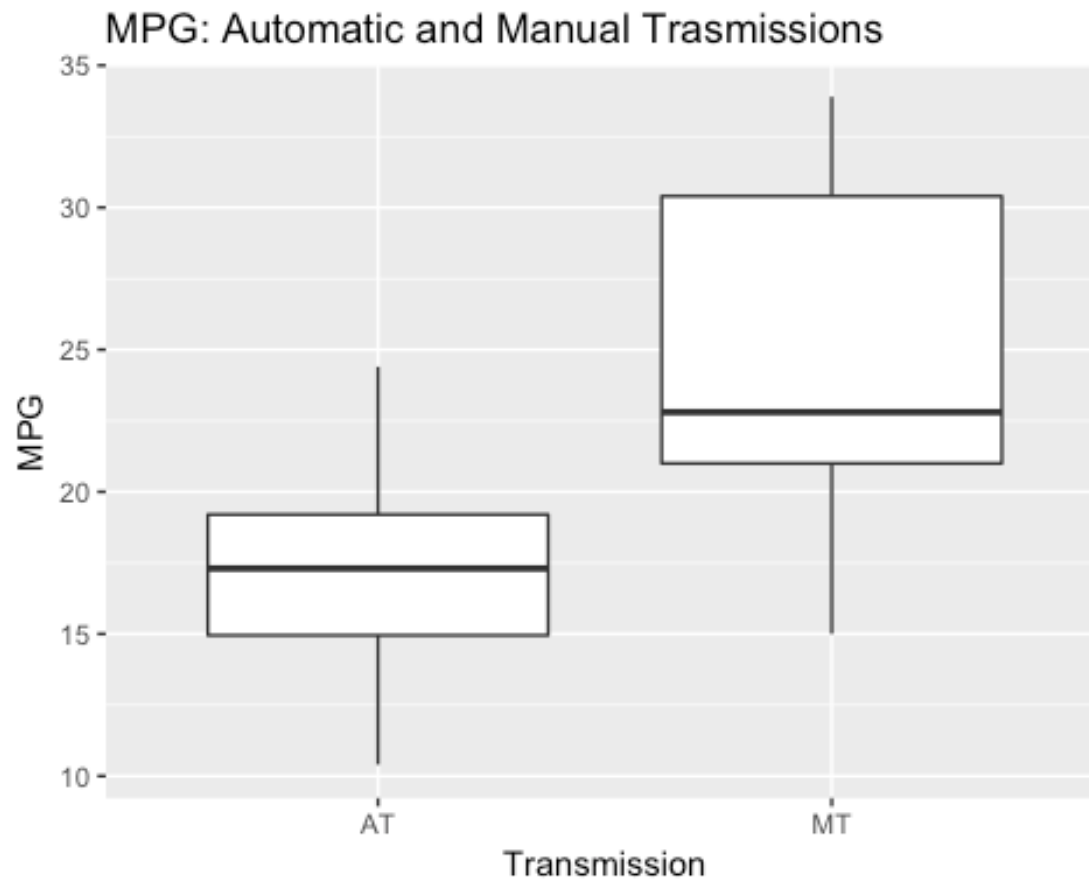
Correlations:

```r
corr <- select(mtcars, mpg,cyl,disp,wt,qsec, am)
pairs(corr, col = 4)
```

Linear Model 1

Illastration mpg for automatic transmisions

```
f1 <-lm(mpg~am, data = mtcars)
summary(f1)

##
## 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 ***
## amMT           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
```

From this linear regression model of mpg against automatic, manual transmission have 7.24 MPG more than automatic transmission. The R^2 value of this model is 0.3598, meaning that it only explains 35.98% of the

Linear Model 2

Using step function.
```
f2 = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(f2)

##
## Call:
## lm(formula = mpg ~ wt + qsec + am, 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
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amMT          2.9358     1.4109   2.081 0.046716 *
## ---
## 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
```

This model uses an algorithm to pick the variables with the most affect on mpg. From the model, the weight, acceleration as well as the transmission affect the mpg of the car the most. Based on a multivariate regression model, a manual transmission cars have better fuel efficiency of 2.94 MPG higher than automatic transmission cars. The adjusted R^2 of the model is 0.834, meaning that 83% of the variance in mpg is do to the model.

ANOVA 2 Models
```
fstep<-lm(mpg~ am + wt + qsec, data = mtcars)
anova(f1, fstep)

## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
```

```
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value indicates that we should reject the null hypothesis that the means from both models are the same. That is, the weight and acceleration of the car have a significant impact on it's MPG.

Conclusion

In conclusion, holding the weight and acceleration of the cars as constant, manual transmission cars offer 2.94 MPG better fuel efficiency.

Model Residuals

```
par(mfrow = c(2,2))
plot(f2, col = 4)
```
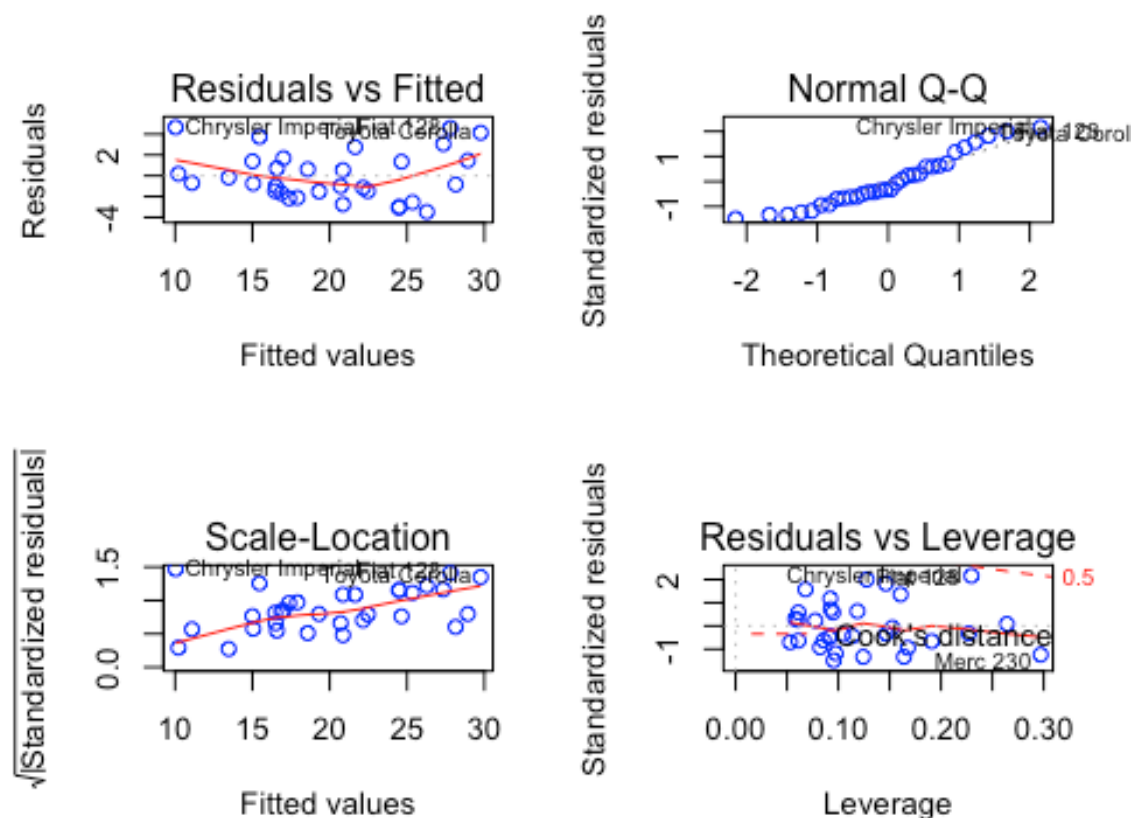

By examining the plot of residuals, we can see that there are a few outliers, but nothing significant that would skew the data.