Week4

Graham Payette

March 30, 2017

Week4 Project Assignment March 30, 2017

Executive summary

For this assignment I performed an exploratory data analysis on the mtcars data set and then looked to find a regression model that best explained the relationship between miles per gallon and type of transmission. (manual versus automatic)

The two questions we were asked to answer were:

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

Based on the analysis the answers are:

1. In the mtcars data set a manual transmission is better for MPG than an automatic transmissions.
2. The cars with manual transmission have a MPG 7.24 higher than the cars with automatic transmission if you just look at the am and mpg data. Model analysis points to a significant relationship between weight and miles per gallon that flattens this advantage to approximately .0236

What?

Like most things...the answer isn't that black and white. Although it is clear that within the mtcars data set manual transmissions show better gas mileage than automatics it is also clear that there are several additional factors that contribute to this with the most significant being weight. Manual transmission cars are generally lighter than equivalent automatic cars. There are several factors (displacement of the engine, number of cylinders in addition to manual/automatic that influence the weight). The above answers still hold true but need to be caveated with the additional information that transmission type is not a universal predictor of MPG.

Through the analysis I tried to find a model that most honestly described the relationship between the overall mtcars data set and fuel economy. With the model I landed on the coefficient for the transmission variable has an estimated value of -0.0236, meaning that a car with manual transmission should have 0.0236 better per gallon fuel economy than a similar car with automatic transmission. The 95% confidence interval for this coefficient

exceeds this value (-3.1848, 3.1376) which means we need more data and analysis to reliably predict MPG base on transmission type.

Approach

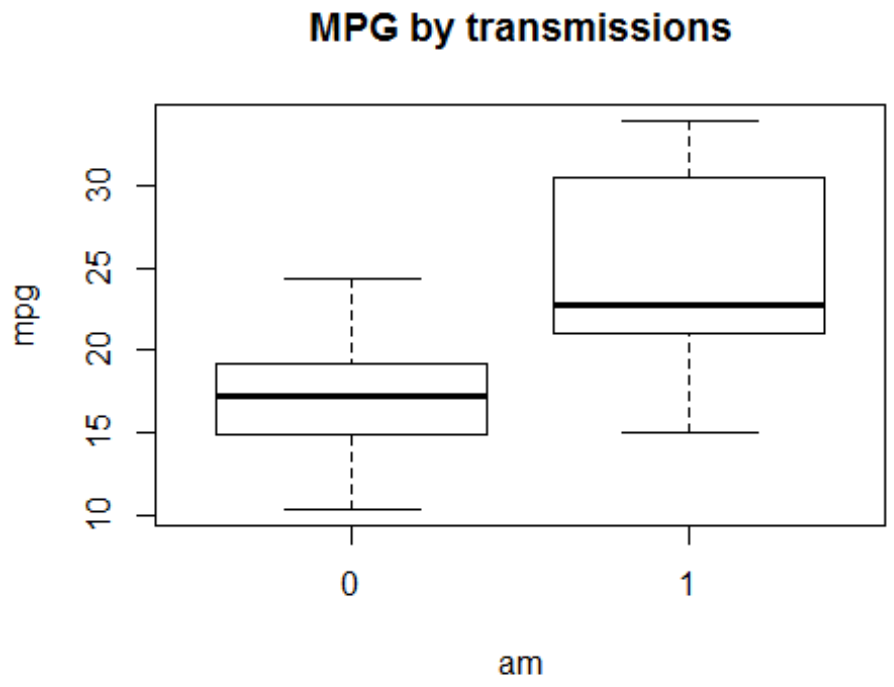
The following details the exploratory analysis, modeling and associated interpretations.

```
head(mtcars)
```

```
##                    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
```

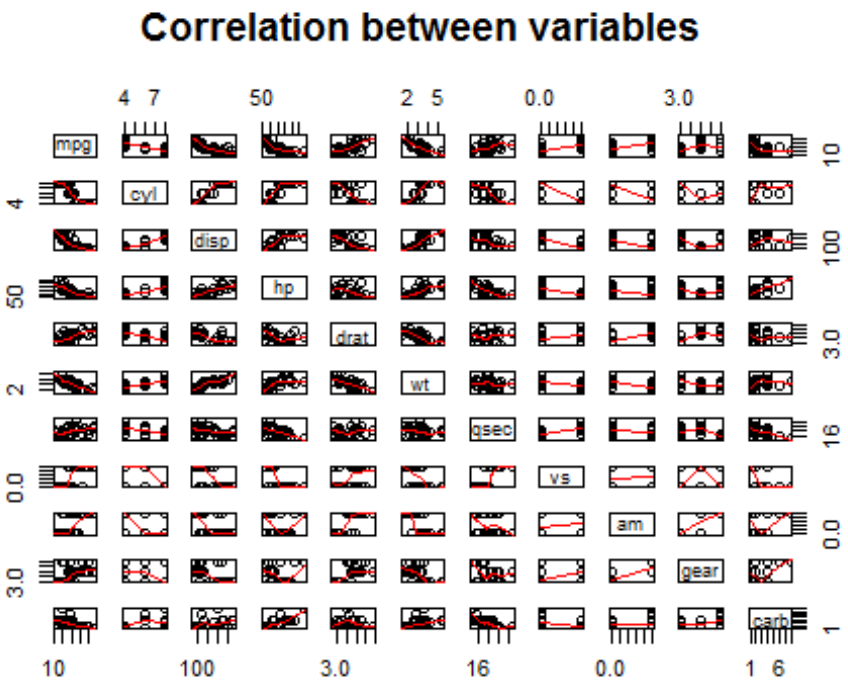
```
#Take a look at the mtcars data set

with(mtcars, {
    plot(as.factor(am), mpg, main="MPG by transmissions", xlab="am",
ylab="mpg")
})
```




```
#The plot shows that the mpg for cars with manual transmissions is better
compared with cars with automatic transmissions.
```

```
require(graphics)
pairs(mtcars,
      main = "Correlation between variables",
      panel=panel.smooth)
```




```
# Explore the correlation between all of the variables in the mtcars data
set.  Note the strong correlation between Wt and MPG

# Generate the basic model and take a look at it
fitam <- lm(mpg ~ am, mtcars)
summary(fitam)$coef

##              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

# Generate the model with all fields and take a look at it
fitall <- lm(mpg ~ ., mtcars)
summary(fitall)$coef

##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
```

```
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871

#Note the significance of WT compared to the other fields. Based on this I am
going to generate a third model containing just AM and WT

fitwt <- lm(mpg ~ am + wt, mtcars)
summary(fitwt)$coef

##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07

#Comparing our three models
anova(fitam, fitall, fitwt)

## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 3: mpg ~ am + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     21 147.49  9    573.40 9.0711 1.779e-05 ***
## 3     29 278.32 -8   -130.83 2.3283   0.05774 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Based on this fitwt looks like the best predictor.  The relation between
auto/manual though is within the error range therefore we can't say AM is a
predictor of fuel economy.
```