Exploring the relationship between a set of variables and miles per gallon (MPG)

Clemens

7/1/2020

Executive Summary

This analysis used the R dataset mtcars to explore the relationship between a set of variables and miles per gallon (MPG).

Through the data analysis presented below we found that in the mtcars dataset automatic transmissions have **1.56 mpg lower** gas mileage than manual transmissions. Further analysis and the backing evidence on the multivariate model selected is included below.

Data Analysis

Exploratory Data Analysis

First we'll load the libraries used for this analysis.

```
invisible({capture.output({
library(dplyr) # above items are silencing the output so don't see all comments about loading dplyr
})})
```

Next we'll load the dataset and take a look at some basic characteristics of mtcars.

```
dim(mtcars)
```

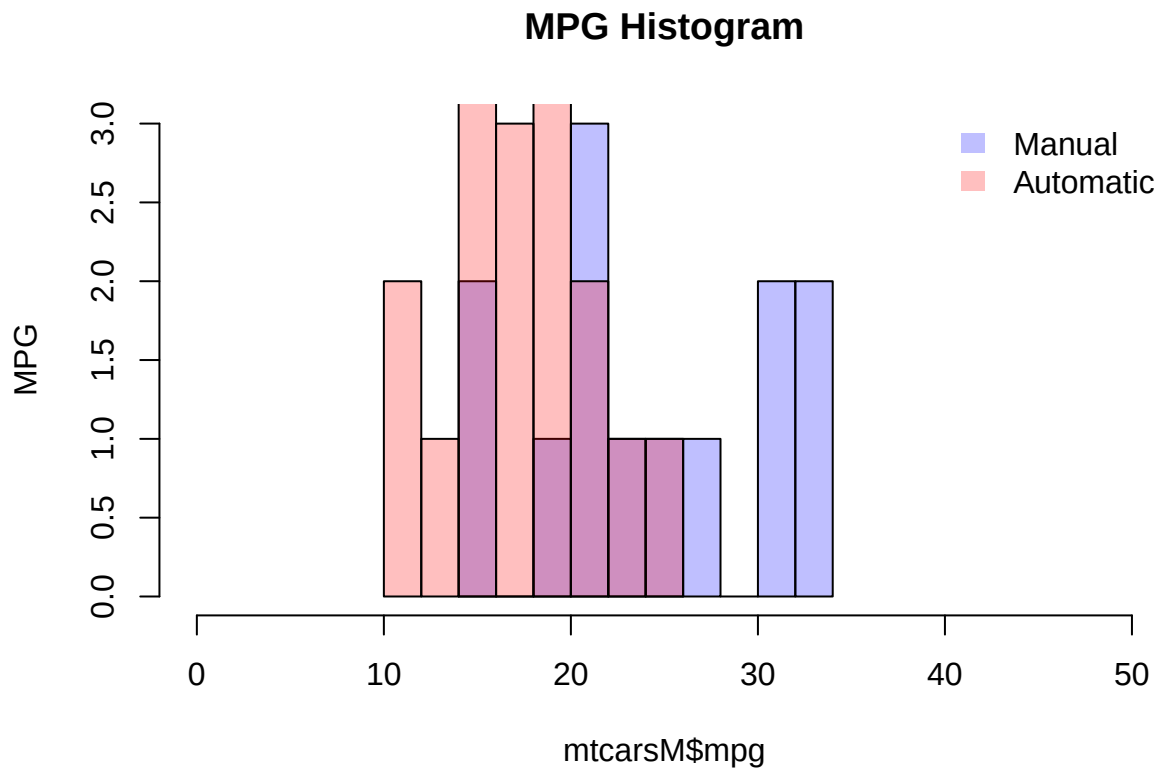
```
## [1] 32 11
```

```
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
```

Trying to get a preliminary idea of the relationships between transmission type and mpg:

```r
mtcars$am <- factor(mtcars$am)
mtcarsM <- filter(mtcars, am == 1)
mtcarsA <- filter(mtcars, am == 0)
hist(mtcarsM$mpg, col=rgb(0,0,1,1/4), xlim=c(0,50), main = "MPG Histogram", ylab = "MPG", breaks = 8)
hist(mtcarsA$mpg, col=rgb(1,0,0,1/4), xlim=c(0,50), add = TRUE, breaks = 8)  # second
legend('topright',c('Manual','Automatic'),
       fill = c(rgb(0,0,1,1/4), rgb(1,0,0,1/4)), bty = 'n',
       border = NA)
```



We can do a quick t.test to check whether there is a significant difference between the effect of automatic and manual transmissions on mpg, which looks likely from the histogram above.

```r
t.test(mtcarsM$mpg,mtcarsA$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcarsM$mpg and mtcarsA$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:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

The confidence interval does not overlap zero and we have a small p-value of .001374, so we can say that there is a significant difference between the two (looking for .05 or lower).

Model Selection and Diagnostics

We'll start with fitting a linear regression model to the data

```
l_mdl <- lm(mpg ~ am, mtcars)
summary(l_mdl)
```

```
##
## 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 ***
## am1            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
```

Because automatic transmissions are denoted by "0" in the mtcars dataset, we see it's coefficient listed as the intercept. We also know that this is the mean mpg of the automatic transmission dataset, **17.147 mpg**. We also see that, on average, manual transmissions have a higher mpg by **7.245 mpg**. At first look these figures seem to have a high signifigance and are more than two standard errors away from 0. However, the R^2 value is 0.36 meaning the model only explains 36% of the variance.

We'll run a multivariate model to see the influence of transmission type when also factoring in other variables, such as car weight.

```
m_mdl <- lm(formula = mpg ~ am + cyl + disp + hp + wt, data = mtcars)
summary(m_mdl)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5952 -1.5864 -0.7157  1.2821  5.5725
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.20280    3.66910  10.412 9.08e-11 ***
```

```
## am1          1.55649    1.44054   1.080  0.28984
## cyl         -1.10638    0.67636  -1.636  0.11393
## disp         0.01226    0.01171   1.047  0.30472
## hp          -0.02796    0.01392  -2.008  0.05510 .
## wt          -3.30262    1.13364  -2.913  0.00726 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 26 degrees of freedom
## Multiple R-squared:  0.8551, Adjusted R-squared:  0.8273
## F-statistic:  30.7 on 5 and 26 DF,  p-value: 4.029e-10
```

We can see with R^2 = .83 that this model explains 83% of the variance. We can also do a quick check comparison with anova.

```
anova(l_mdl, m_mdl)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + disp + hp + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 163.12  4    557.78 22.226 4.507e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of **4.507e-08** we confirm the multivariate model is significantly better than the single linear model.

Looking back at the coefficients of the multivariate model, it tells us that manual transmissions ("am1") have **1.56 mpg higher** gas mileage than automatic transmissions. Furhter diagnostic plots, including residuals, are included in the appendix below.

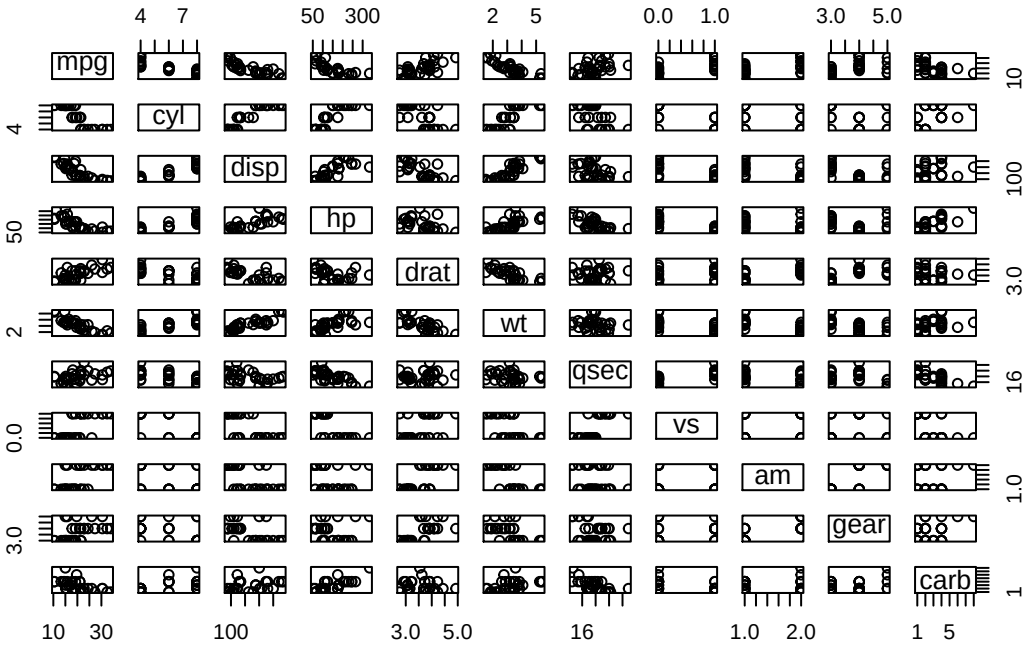
Conclusions

In conclusion, the analysis on mtcars shows that automatic transmissions have better gas mileage than manual transmissions. Strictly looking at transmission types, automatic transmissions in mtcars, on average, have a lower gas mileage by 7.245 mpg. Fine tuning our model selection and accounting for other variables, such as car weight, we foud a significant result (our model explaining 83% of the variance in mpg) that **automatic transmissions have 1.56 lower mpg gas mileage than manual transmissions**.

Appendix

Pairs Plot to look variable influence on mpg

```
pairs(mpg ~ ., mtcars)
```



Multivariate regression diagnostic plots

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