Fuel Effecincy Evaluation

Seth Dobrin

February 16th, 2016

1 Executive Summary

This study eplores the relationship of fuel effeciancy (measure as miles per gallon) and transmission type utilizing regressions analysis. In this study we will look at the effect of transmission type on MPG performance and answer the question: Is an automatic or manual transmission better for MPG? Additionally, we will quantify the MPG difference between automatic and manual transmissions. As part of this analysis we will explore other factors that may contribute to MPG performance #Data Processing & Eploration Load the data and perfom basic exploration

```
data(mtcars)
M <- cor(mtcars)
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))  #convert 'am' to factor
mtcars$names <- as.factor(rownames(mtcars))
n <- length(mtcars$mpg)
alpha <- 0.05
head(mtcars, 5)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs        am gear
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0    Manual    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0    Manual    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1    Manual    4
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1 Automatic    3
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0 Automatic    3
##                   carb             names
## Mazda RX4            4         Mazda RX4
## Mazda RX4 Wag        4     Mazda RX4 Wag
## Datsun 710           1        Datsun 710
## Hornet 4 Drive       1    Hornet 4 Drive
## Hornet Sportabout    2 Hornet Sportabout
```

2 Eploratory Analysis

Figure 1 examines the correlations accross the data set. If you use this to visually examine which factors are correlated with a fuel consumption (mpg) you can see that the strongest correlations are number of cylinders (cyl), displacement (disp), horsepower (hp) and weight (wt). Transmisison type has a lesser correlation, but is none the less still correlated to mpg and will be considered as requested dispite the weaker correlation. Figure 2 examines the varriance between mpg and transmission type by creating a boxplot that demonstrates a slight significant increase in MPG with a manual transmission. This is futher confirmed by performing a t-test which yields a p-value of 0.0014 (Table 1). ##Paired t-test Table 1 shows the output of the t-test with significance demonstrated for an increase in fuel effeciancy accosiated with transmission type with a p-value of 0.01374.

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

2.1 Regression Analysis

Simple model of miles per galon (mpg) and transmission type (am) (Table 2)

```
mtBaseModel <- lm(mpg ~ am, data = mtcars)
baseResid <- summary(mtBaseModel)$sigma
```

Figure 1 shows the factors correlated to mpg. These factors are used to crerate multiple regression models (Table 3).

```
mtMultiModel <- lm(mpg ~ am + cyl + disp + wt, data = mtcars)
```

Examine the variance accross the diferent models using an ANOVA (Table 4)

```
model1 <- lm(mpg ~ am, data = mtcars)
model2 <- lm(mpg ~ am + cyl, data = mtcars)
model3 <- update(model1, mpg ~ am + cyl + disp)
model4 <- update(model1, mpg ~ am + cyl + disp + wt)
sig <- anova(model1, model2, model3, model4)
```

Table 4 show the output of the ANOVA. Models 2 and 4 are highly sigificant with p-vales of 1.264e-08 and 0.05468 respectively

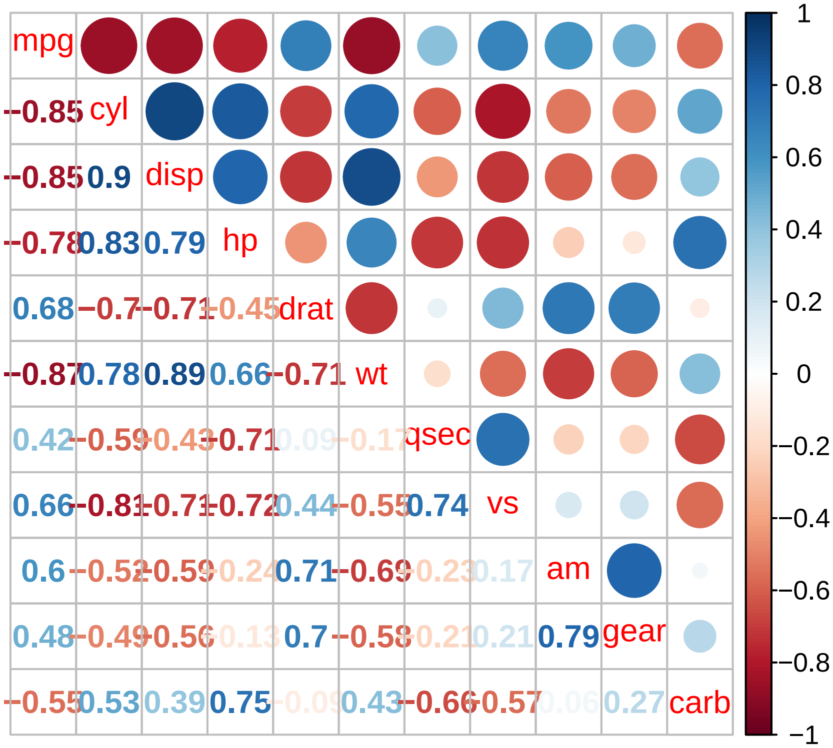
3 Conclusions

Individually each of the factors effect the fueld effeicency of the cars in the data set. The ANOVA demonstrates the significance of the multifactor model. This multifactor model enhances the predictability. The muli model provides a much better fit as demonstarted by th R^2 values. The base model has an R^2 of 0.3598 and the multi model has an R^2 value of 0.8327 as laid out in Tables 2 & 3 respectively. Table 4 shows the significance of each model used in the multi factor regression analaysis as determined by an analysis of variance (ANOVA). Models 1, 2 and 4 show high significance based on the inital paired t-tes for models 1 and the ANOVA for models 2 and 4. A manual transmission improves fuel effcency, however the combination of an automatic transmisison, 4 cylindar engine, low displaement and low weight have an even more significant impact on fuel effciency.

4 Appendix

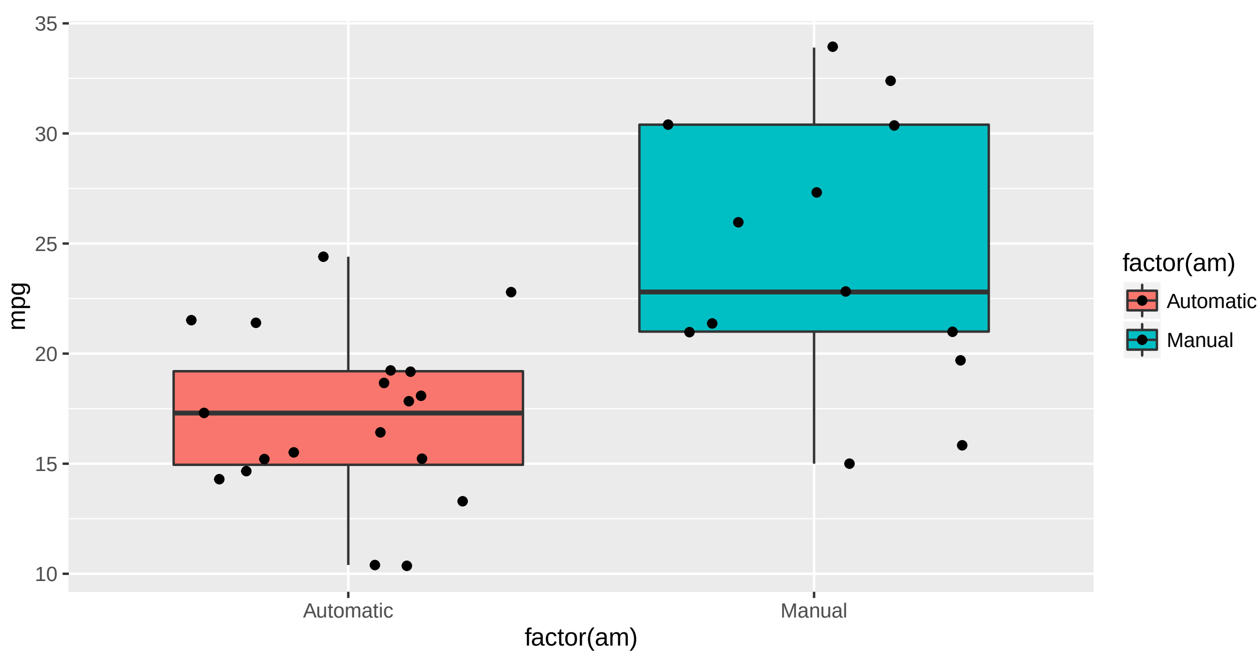
4.1 Figure 1. Correlation Plot.

```
library(corrplot)
corrplot.mixed(M, sig.level = 0.05, tl.pos = "d", tl.srt = 60)
```



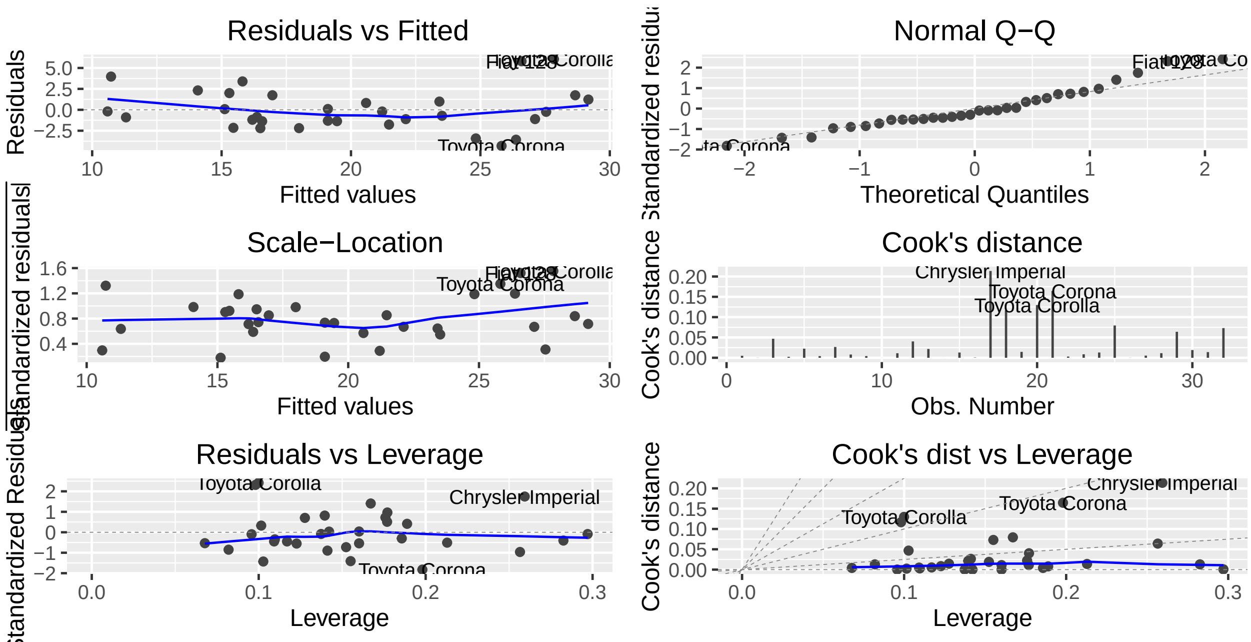
4.2 Figure 2. Boxplot of miles per galon by transmission type

```
library(ggplot2)
qplot(factor(am), mpg, data = mtcars, geom = c("boxplot", "jitter"), fill = factor(am))
```



4.3 Figure 3. Diagnostic Plots (includes residual)

```r
library(ggfortify)
autoplot(lm(mpg ~ am + cyl + disp + wt, data = mtcars), label.size = 3, ncol = 2,
    which = 1:6)
```



4.4 Table 1: Summary of t-test

```r
library(pander)
pandoc.table(glance(t.test(mpg ~ am, data = mtcars)), style = "simple")
```

```
##
##
##  estimate   estimate1   estimate2   statistic   p.value   parameter   conf.low
## ---------- ----------- ----------- ----------- --------- ----------- ----------
##   -7.245      17.15       24.39      -3.767    0.001374     18.33      -11.28
##
## Table: Table continues below
##
##
##
##  conf.high
## -----------
##    -3.21
```

4.5 Table 2: Summary of base model

```r
pandoc.table(glance(mtBaseModel), style = "simple")
```

```
##
##
##  r.squared   adj.r.squared   sigma   statistic   p.value   df   logLik   AIC
## ----------- --------------- ------- ----------- --------- ---- -------- -----
##   0.3598        0.3385       4.902     16.86    0.000285   2    -95.24  196.5
##
## Table: Table continues below
##
##
##
##  BIC   deviance   df.residual
## ----- ---------- -------------
## 200.9   720.9         30
```

4.6 Table 3: Summary of multi-factor model

```r
pandoc.table(glance(mtMultiModel), style = "simple")
```

```
##
##
##  r.squared   adj.r.squared   sigma   statistic   p.value   df   logLik   AIC
## ----------- --------------- ------- ----------- --------- ---- -------- -----
##   0.8327        0.8079       2.642     33.59    4.038e-10  5    -73.77  159.5
##
## Table: Table continues below
##
##
##
##  BIC   deviance   df.residual
## ----- ---------- -------------
## 168.3   188.4         27
```

4.7 Table 4: Summary of anaysis of variance

```r
pandoc.table(tidy(sig), style = "simple")
```

```
##
##
##  res.df   rss   df   sumsq   statistic   p.value
## -------- ----- ---- ------- ----------- ---------
##    30    720.9  NA    NA        NA          NA
##    29    271.4  1    449.5     64.41    1.264e-08
##    28    252.1  1    19.28     2.763      0.108
##    27    188.4  1    63.66     9.121    0.005468
```