Effect of auto/manual transmission on MPG in motor cars

Amelia

1 March 2017

Executive Summary

This report investigates if automatic or manual transmission better for MPG in motor cars, and quantify the MPG difference between automatic and manual transmissions. Models variables are selected based on variables with the strongest correlation to MPG and tested with a nested model. Based on the results, a model with the following variables: am, wt, cyl was selected. Based on the model selected, changing from auto to manual increase the millage by 0.1765 miles/gallon. The 95% confidence interval based on model 3 is prediction + 5.27907 or -5.27907.

Model Selection

From the boxplot, the manual cars have a better miles/gallon record than auto cars. (see appendix Figure 1)

For this project, 3 linear models will be compared. The selection of these variable are based on the variables with the strongest correlation. (see appendix Figure 2)

Model 1: Response variable: miles/gallon, Predictor variable: auto/manual transmission as factor only Model 2: Response variable: miles/gallon, Predictor variable: auto/manual transmission as factor and wt. Model 3: Response variable: miles/gallon, Predictor variable: auto/manual transmission as factor, weight and cylinder.

Using nested model testing, we test if the additional variable adds information to the model.

```
mod1<-lm(mpg~ factor(am),data = mtcars)
mod2<-lm(mpg~ factor(am)+wt,data = mtcars)
mod3<-lm(mpg~ factor(am)+wt+cyl,data = mtcars)
anova(mod1,mod2,mod3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + cyl
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 64.864 9.051e-09 ***
## 3     28 191.05  1     87.27 12.791  0.001292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results shows that we can reject the null hypothesis and the addiional factor wt adds informtion to the model. Hence all predictor variables are significant. Model 3 is selected over model 1 and 2.

```
summary(mod3)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + cyl, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.4179     2.6415  14.923 7.42e-15 ***
## factor(am)1   0.1765     1.3045   0.135  0.89334
## wt           -3.1251     0.9109  -3.431  0.00189 **
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

```
# confidence intervals
mod3coef<- coef(summary(mod3))
error<- sum(mod3coef[,2])
```

Conclusion

Based on the model, changing from auto to manual increase the millage by 0.1765 miles/gallon. The 95% confidence interval based on model 3 is prediction + 5.27907 or -5.27907. See appendix Figure 3 for residual analysis

Appendix

EDA

Boxplot

```
boxplot(mpg~am, data = mtcars, xlab = "0=auto, 1=manual", ylab = "miles/gallon")
```

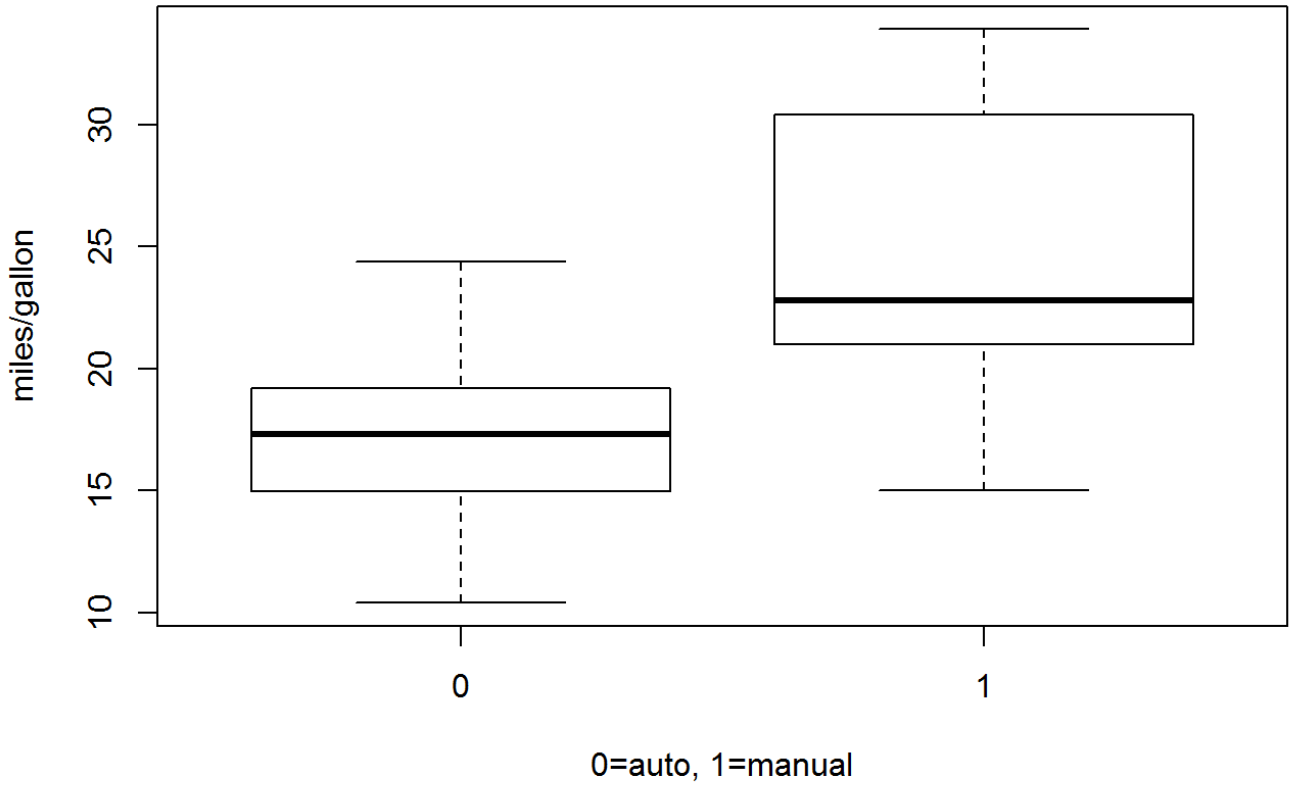

Figure 1: Manual cars have a better miles/gallon record than auto cars.

Correlation plot

```
library(ggcorrplot)
corr <- round(cor(mtcars), 1)
head(corr[, 1:6])
```

```
##       mpg  cyl disp   hp drat   wt
## mpg   1.0 -0.9 -0.8 -0.8  0.7 -0.9
## cyl  -0.9  1.0  0.9  0.8 -0.7  0.8
## disp -0.8  0.9  1.0  0.8 -0.7  0.9
## hp   -0.8  0.8  0.8  1.0 -0.4  0.7
## drat  0.7 -0.7 -0.7 -0.4  1.0 -0.7
## wt   -0.9  0.8  0.9  0.7 -0.7  1.0
```

```
p.mat <- cor_pmat(mtcars)
head(p.mat[, 1:4])
```

```
##                mpg          cyl         disp           hp
## mpg  0.000000e+00 6.112687e-10 9.380327e-10 1.787835e-07
## cyl  6.112687e-10 0.000000e+00 1.802838e-12 3.477861e-09
## disp 9.380327e-10 1.802838e-12 0.000000e+00 7.142679e-08
## hp   1.787835e-07 3.477861e-09 7.142679e-08 0.000000e+00
## drat 1.776240e-05 8.244636e-06 5.282022e-06 9.988772e-03
## wt   1.293959e-10 1.217567e-07 1.222320e-11 4.145827e-05
```

```
ggcorrplot(corr)
```

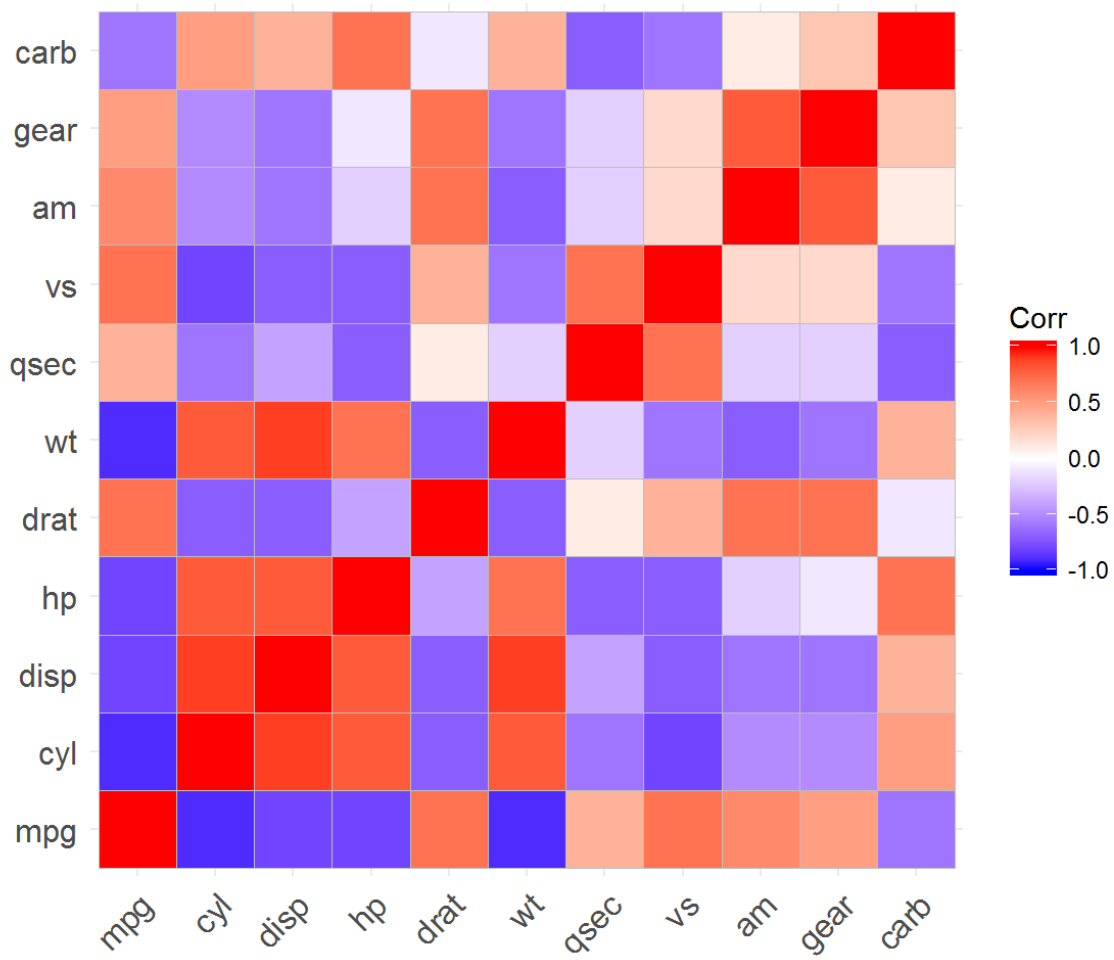

Figure 2: correlation plot shows that the strongest correlations to mpg is wt and cyl, and the correlation is negative.

Residual analysis

```
par(mfrow=c(2,2))
plot(mod3)
```

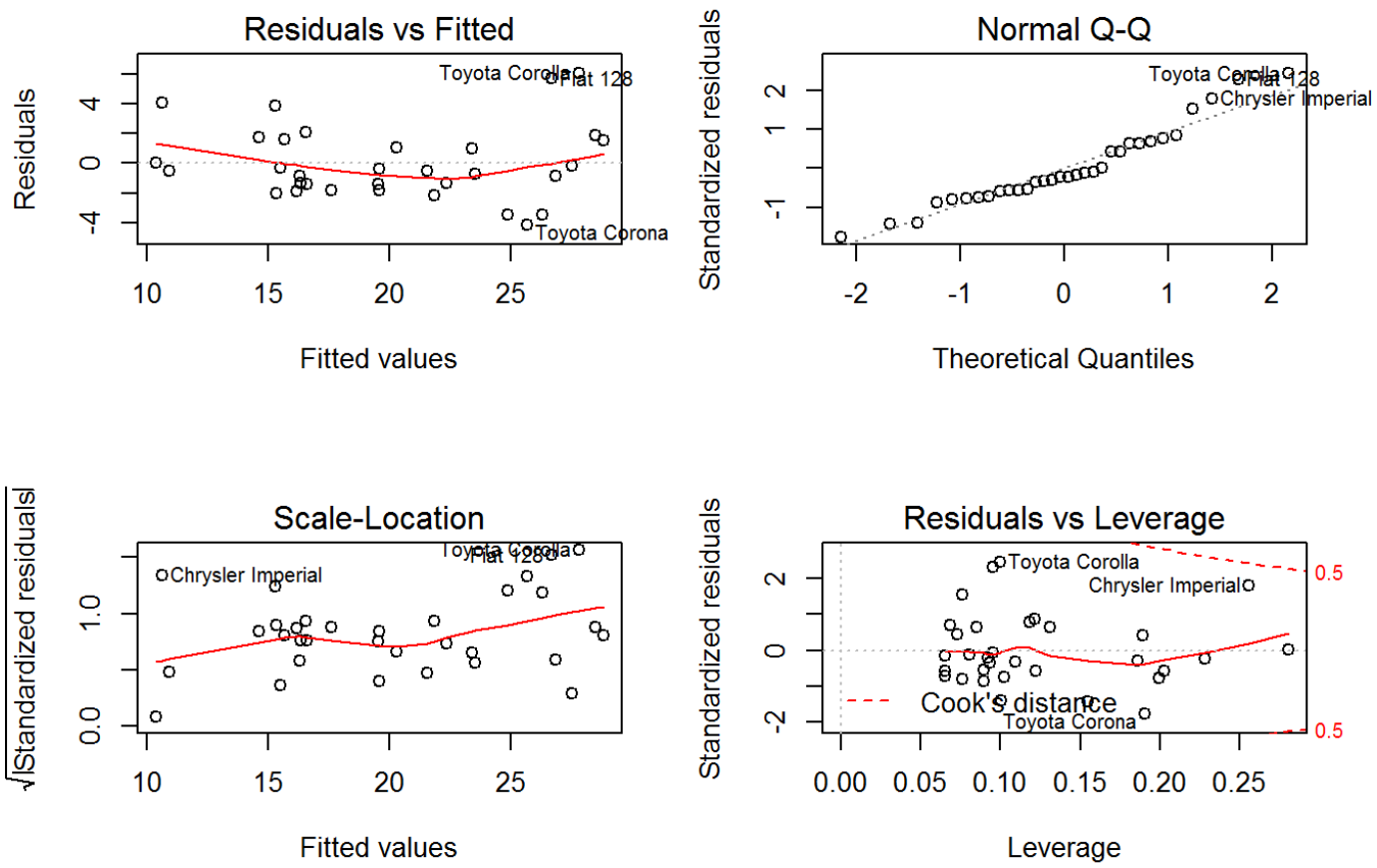

Figure 3: There are 2 points with higher residual at highest mpg. The the remaining residuals are well spread across the different mpg scale and thus this model is likely to be a good fit.