Is an automatic or manual transmission better for MPG

Sayef Ishauqe

October 21, 2017

Executive Summary

This paper looks into the R standard data `mtcars' to answer the question **Is an automatic or manual transmission better for MPG?** and if so, **what is difference in MPG for automatic vs manual transmission**

Method

The analysis concentrates on identifying the best possible multi-variable regression model to calcualte mpg and then interpret the infulence onf transmission being manual or automatic in the model.

Analyze the correlation of variables

- A ggpairs chart has been generated (***Figure 1***) to see the correlation of variables with mpg.
    - `mpg` has high correlation with weight (`wt`), cylinders (`cyl`) and displacement (`disp`)
    - `cyl` and `disp` has high correlation within them. Considering the audience of the magazine `cyl` could be a simiplar predictor to communicate
    - Trasmission, automated or manual (`am`) seems doesn't have high coorelation with `mpg` .
    - So, developing a multi-variable regression model with minimum number of variable is the best way to move. Once, the model is achieved transmission variable can be introduced to see the impact (for inference / hypothesis).

Construct and compare models

- Fit a linear regression model combination of variables as identified above
    - Test the model with log of wight as well to see it is a better fit. This is because it is not logical to calculate mpg for a car with 0 wight.

```
fit1<-lm(mpg~wt,data=mtcars)
fit2<-lm(mpg~wt+factor(am),data=mtcars)
fit3<-lm(mpg~wt+factor(cyl),data=mtcars)
fit4<-lm(mpg~wt+factor(am)+factor(cyl),data=mtcars)
fit5<-lm(mpg~I(log(wt))+factor(am)+factor(cyl),data=mtcars)
anova(fit1,fit2,fit3,fit4,fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + factor(am)
## Model 3: mpg ~ wt + factor(cyl)
## Model 4: mpg ~ wt + factor(am) + factor(cyl)
## Model 5: mpg ~ I(log(wt)) + factor(am) + factor(cyl)
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 278.32
## 2     29 278.32  1     0.002  0.0003 0.9856271
## 3     28 183.06  1    95.261 14.0573 0.0008557 ***
## 4     27 182.97  1     0.090  0.0133 0.9089474
## 5     27 156.21  0    26.762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model

mpg = 37.02 -12.33 x log(wt) - [0.7 for manual transmission] - [3.03 for 6 cyl or, 4.91 8 cyl]

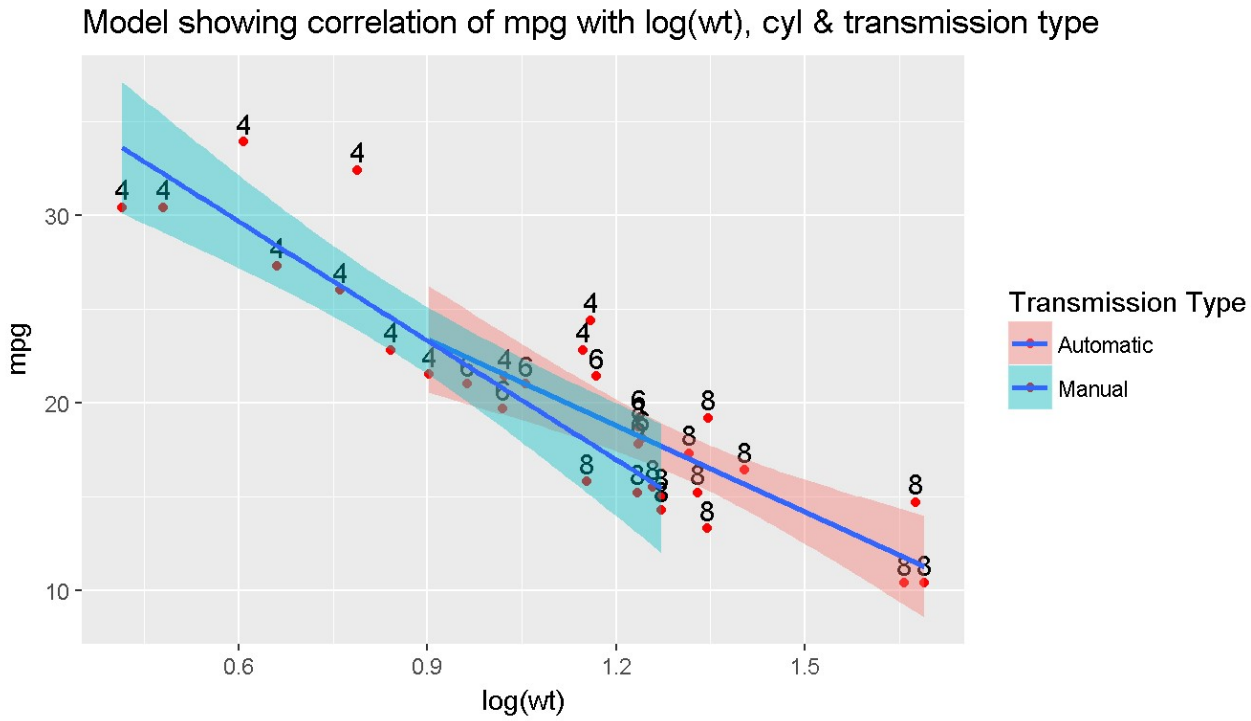
```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)   37.024076   3.008701 12.3056666 1.392662e-12
## I(log(wt))   -12.332869   2.850625 -4.3263745 1.857436e-04
## factor(am)1   -0.704919   1.265183 -0.5571677 5.820027e-01
## factor(cyl)6  -3.028674   1.402497 -2.1594865 3.985860e-02
## factor(cyl)8  -4.914960   1.646150 -2.9857295 5.950103e-03
```

```
##                   2.5 %     97.5 %
## (Intercept)   30.850731 43.1974210
## I(log(wt))   -18.181868 -6.4838708
## factor(am)1   -3.300860  1.8910218
## factor(cyl)6  -5.906361 -0.1509873
## factor(cyl)8  -8.292581 -1.5373382
```

- The model also indicates that automatics transmission vehicles (`am==0`) produces slightly higher mpg (0.7) than that of manual transmission vehicle. However, Further looking into the confidence interval of the cofficients, the 95% interval of transimssion variable `am` included 0. So, the hypothesis that this variable has no impact on milage can not be rejected.

Conclusion

The model shows that millage is mostly correlated with weight and than number of cylinders of a car.



While in preliminary anaysis, it showed that manual transmission provides higher mpg (***Figure 2***), the model has produced high residuals. A better fit model shows that transmission type doesn't have significant impact on mpg. Analysis of residuals is shown in ***Figure 3***. This model is still not safe from high residuals which can be further analyzed to refine the multi-variable regressional model.

Appendix

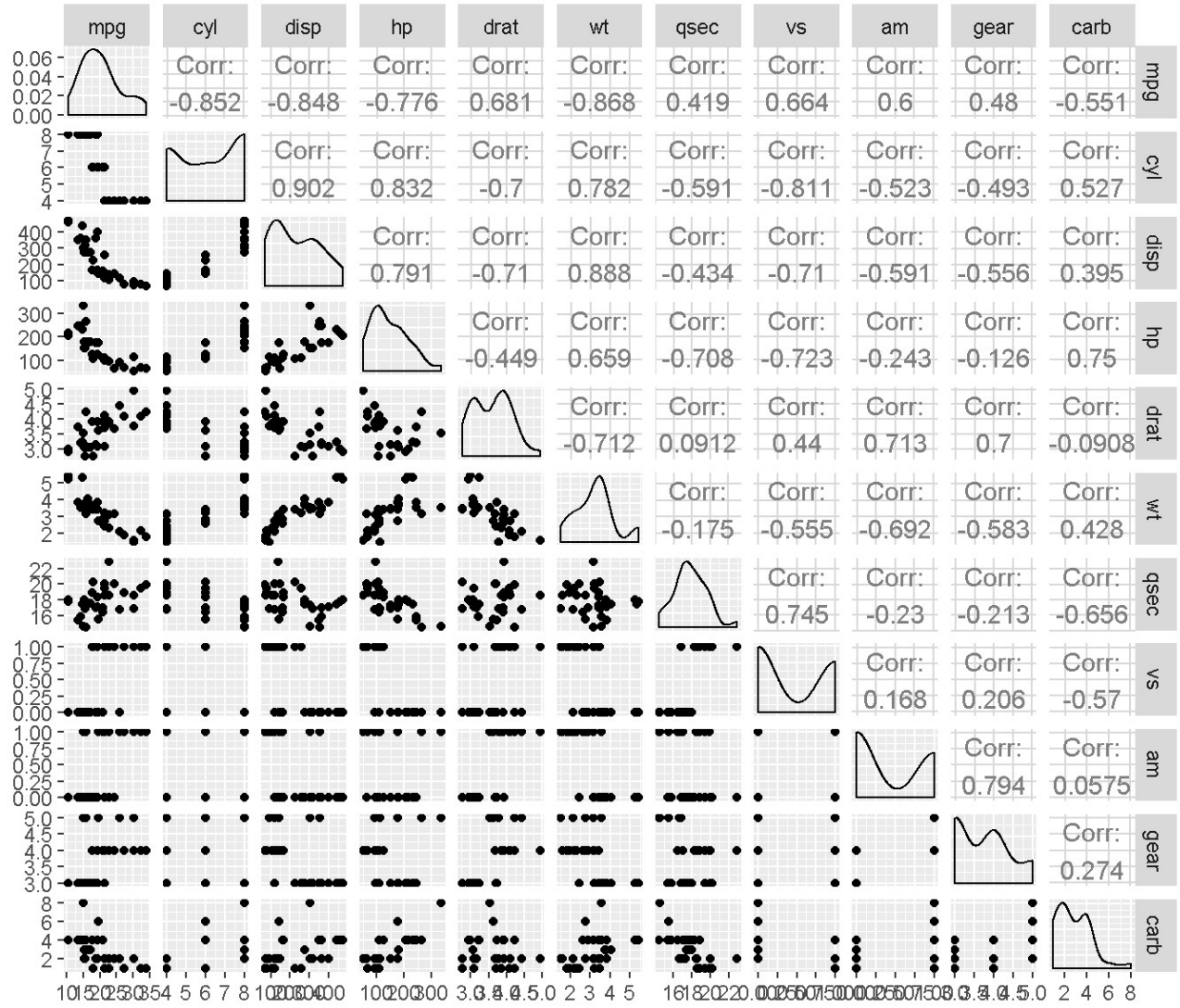


Figure 1: Pairs plot of all variables in mtcars

```
##
## Call:
## lm(formula = mpg ~ factor(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 ***
## factor(am)1    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
```

Figure 2: correlation of transmission type with mpg

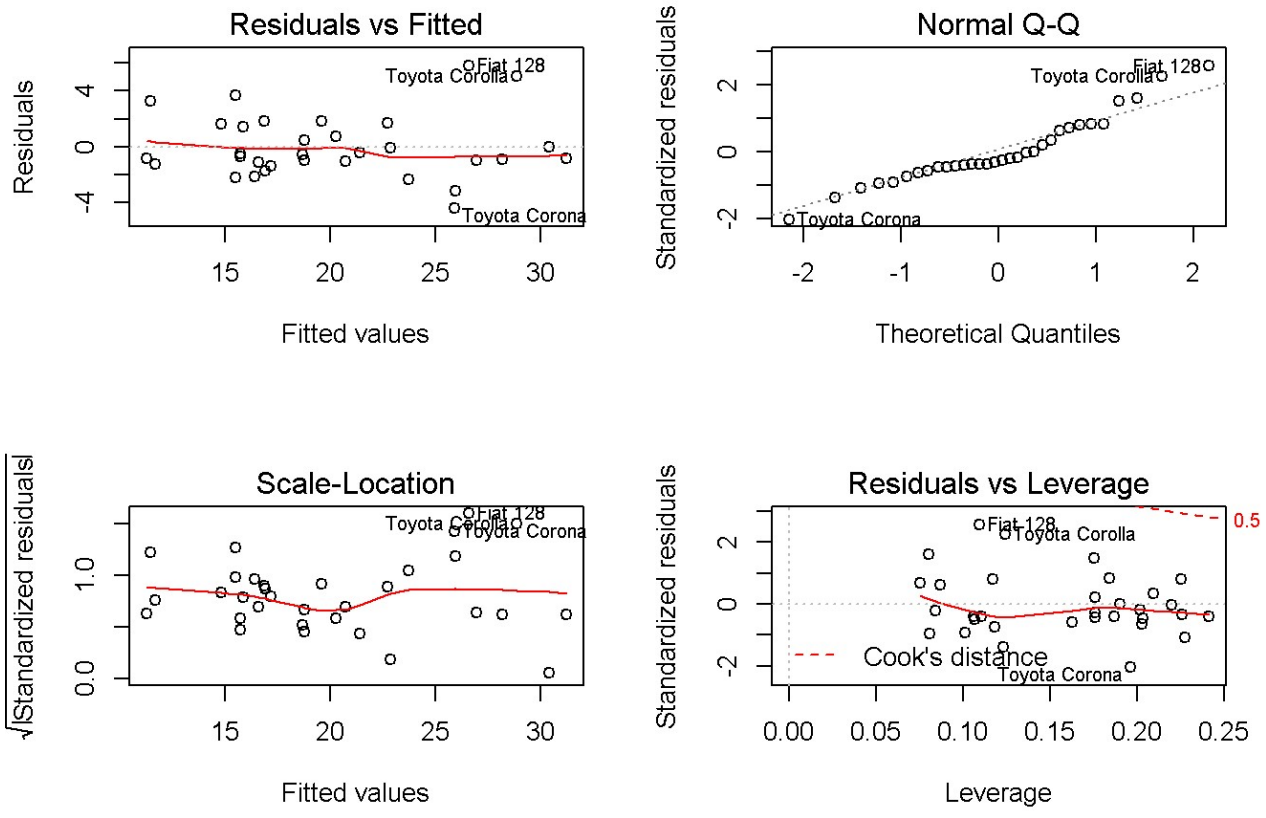


Figure 3: Residuals of the suggested model