Effect of Transmission on Fuel Efficiency

Keh-Harng Feng

March 27, 2017

1 Synopsis

The relationship between fuel efficiency and car transmission type is studied using the `mtcars` dataset. A model is built using MLR and regression analysis is carried out. Hypothesis tests suggest that while weight and horsepower are significant factors in fuel efficiency, transmission type is not.

2 Introduction

The torque and speed output of an internal combustion engine is controlled using a set of gears. Originally, vehicles require the drivers to manually shift the gears during operation according to the road condition and speed of the vehicle. The advent of automatic transmission allows motor vehicles to automate the gear shifting process without direct human input. This report aims to discover the potential effect of transmission type on vehicle fuel efficiency.

3 Data Description & Manipulation

Motor vehicle road test performance is obtained from the mtcars dataset. The origin of the data and a list of variables contained can be found on its R documentation page. The outcome variable of primary interest is `mpg` which measures fuel efficiency. The predictor of interest is `am`, which categorizes transmission type. However other variables may or may not have an effect on fuel efficiency as well. To allow logical interpretations during regression analysis, the following variables are identified as potential non-categorical predictors. They are then centered and renamed:

- hpc <- centered hp
- wtc <- centered weight
- dratc <- centered rear axle ratio

The am variable is also transformed into a factor variable amf with the proper labels 'auto' and 'manual'.

4 Exploratory Data Analysis

Boxplots showing mpg vs transmission type grouped by the categorical predictors are shown in Figure 1. Besides the cases when no data is available for a particular transmission type, all groups show higher mpg for manual transmission. It is therefore safe to focus on possible association between transmission type and non-categorical predictors.

Figure 2 shows how various non-categorical predictors are associated with transmission type. Without carrying out formal hypothesis tests, there seems to be non-trivial differences between automatic and manual transmission. Whether these are confounding variables will be analyzed using multivariate linear regression.

5 Regression Analysis

5.1 Model Selection

Since amf is a two-level factor coded with 0 and 1, it can be directly used as a predictor in MLR. Nested MLR fits are generated with increasing model complexities. ANOVA-F test is then carried out (see appendix).

F-test P values suggest that the inclusion of weight and horsepower is justified while the rest are not. Transmission type, weight and horsepower are therefore selected to be the relevant predictors (ie: fit2). The residual plot is shown in Figure 3.

It can be seen that residuals vs fitted never deviate to more than 2 standard units away from 0. The QQ plot shows that the residual conforms to the normal distribution pretty well although at extreme values it seems to become a bit too fat. Scale-Location vs predictors plot shows a relatively horizontal line so the variance of residual is pretty constant (satisfies homoscedasticity) and finally the points that are furthest away from 0 have pretty low leverage so the fit should be fairly representative of the sample. Overall, the selection seems to be good.

5.2 Model Interpretation

The summary for the selected model are shown below.

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.244      0.717  26.845    0.000
## amfmanual      2.084      1.376   1.514    0.141
## wtc           -2.879      0.905  -3.181    0.004
## hpc           -0.037      0.010  -3.902    0.001
```

Since amf is a factor, the "auto"" level or amf = 0 is used as the reference point. The magnitudes of the above estimates can be interpreted as follows:

- Intercept: A car with average weight and horsepower and automatic transmission can drive 19.24 miles per gallon.
- amfmanual: If a manufacturer produces a manual transmission model of an automatic transmission car, holding things such as weight and horsepower constant, the car can now drive 2.08 miles further than before per gallon.
- wtc: If one is to increase the weight of the car by 1000 lbs without changing anything else, the car will drive 2.879 miles less on one gallon of gas than before.
- hpc: If one is to increase the engine power by 1 hp, the car will drive 0.037 miles less on one gallon of gas than before.

Notice that although the intention is to study the effect of transmission type on fuel efficiency, the estimate for the coefficient associated with transmission type (amfmanual) has a T-test P value of 0.141. This is quite a bit larger than the traditional cutoff value of 0.05. Thus the estimated transmission type coefficient is not significantly larger than what is assumed in the null hypothesis, 0.

6 Conclusion

Using MLR analysis a model is constructed to reflect the association of fuel efficiency, mpg with transmission type, horsepower and weight. Residual analysis shows that the model is a good fit. However, T-test on the model coefficients reveals that the null hypothesis for the transmission type coefficient cannot be rejected. The evidence therefore suggests that the type of transmission **does not** have significant effect on fuel efficiency. Weight and horsepower seem to have much more significance in this regard.

7 Appendix

7.1 Anova Test For Model Selection

```
fit0 <- lm(mpg ~ amf, data = mtcars)

fit1 <- update(fit0, mpg ~ amf + wtc)

fit2 <- update(fit1, mpg ~ amf + wtc + hpc)

fit3 <- update(fit2, mpg ~ amf + wtc + hpc + dratc)

fit4 <- update(fit3, mpg ~ amf + wtc + hpc + dratc + qsec)

fit5 <- update(fit4, mpg ~ amf + wtc + hpc + dratc + qsec + disp)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ amf
## Model 2: mpg ~ amf + wtc
## Model 3: mpg ~ amf + wtc + hpc
## Model 4: mpg ~ amf + wtc + hpc + dratc
## Model 5: mpg ~ amf + wtc + hpc + dratc + qsec
## Model 6: mpg ~ amf + wtc + hpc + dratc + qsec + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 73.7170 6.325e-09 ***
## 3     28 180.29  1     98.03 16.3280 0.0004461 ***
## 4     27 176.96  1      3.33  0.5540 0.4636198
## 5     26 158.64  1     18.33  3.0525 0.0928895 .
## 6     25 150.09  1      8.55  1.4233 0.2440542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7.2 Figures

```
## [1] 2
```

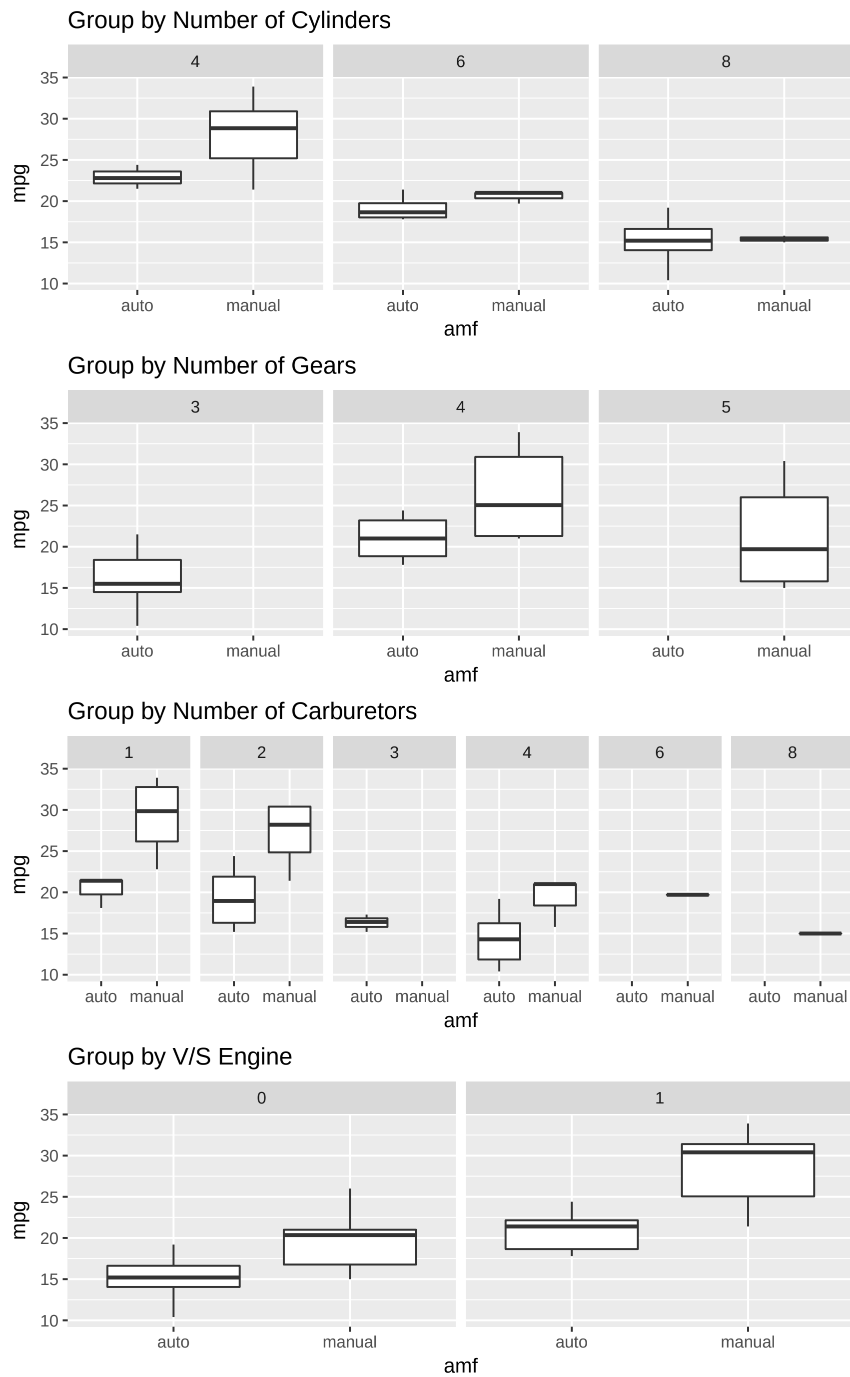

Figure 1: Potential Categorical Predictors Grouped by Transmission Type

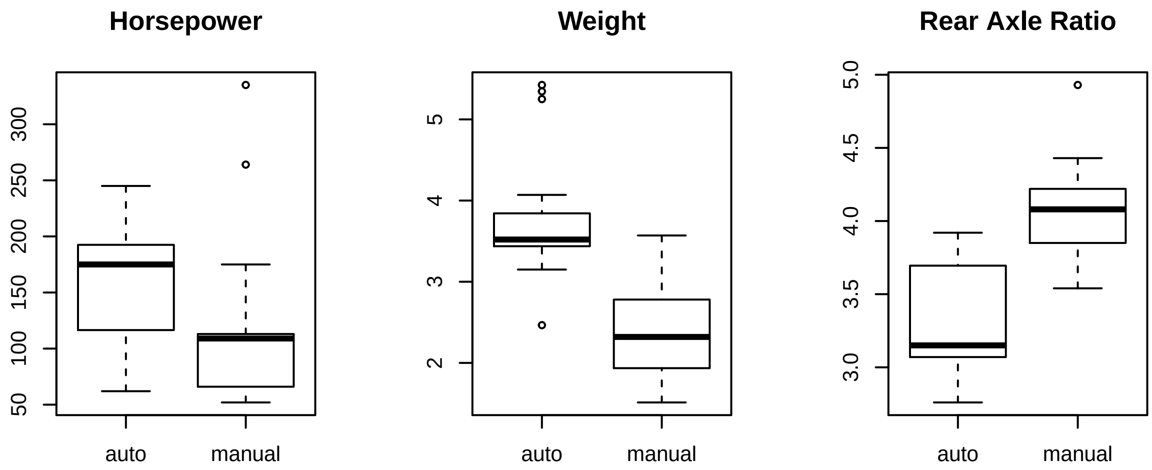

Figure 2: Potential Non-Categorical Predictors Grouped by Transmission Type

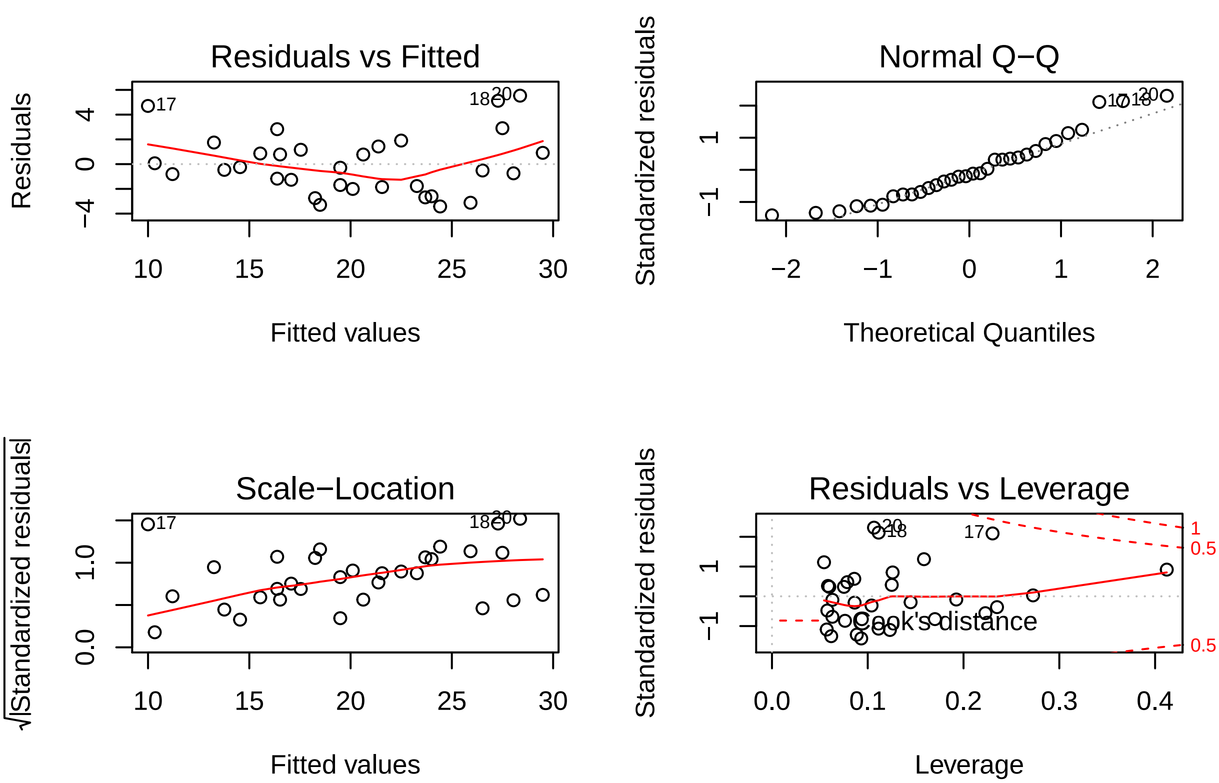

Figure 3: Residual Analysis