Motor Trend Analysis

Jason Murray

2017-02-14

Summary

In this analysis we will be using the Motor Trend data from 1974 that we are all familiar with by now to explore a few questions.

- Is an automatic or manual transmission better for MPG?
- Quantify the difference between automatic and manual transmissions.

Data Load and Exploration

Required libraries

```
library(ggplot2)
```

First let's load up the data

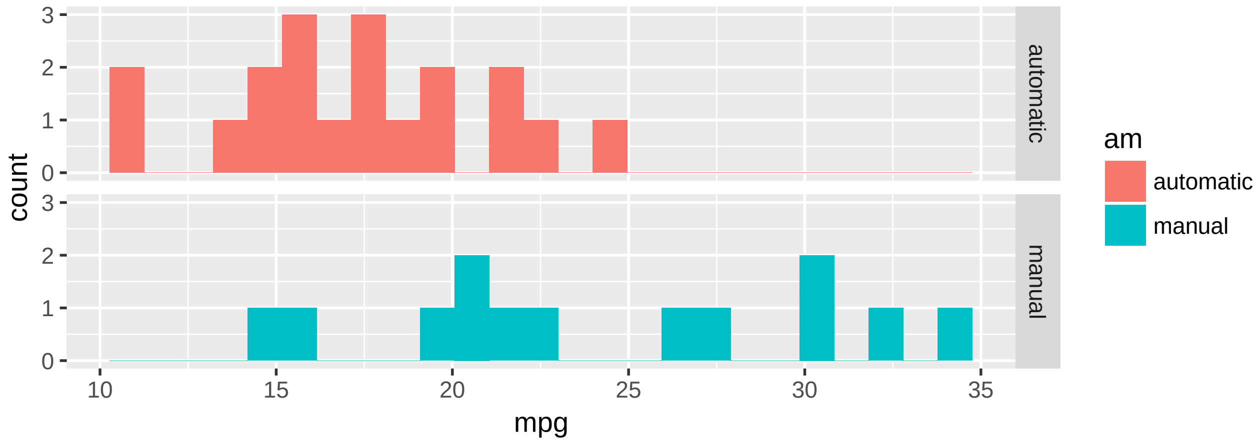
```
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

For some quick initial exploration a histogram of MPG for automatic and manual transmissions might be interesting

```
#Change am to a factor variable and names for the levels
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("automatic", "manual")

ggplot(data=mtcars) +
    geom_histogram(aes(x=mpg, fill=am), bins = 25 ) +
    facet_grid(am ~ .)
```



So initially it looks like manual transmissions may lead to better MPG, but let's dig deeper there could be a lot of other factors influencing this.

Building some models to compare

From the data description. We have the following variables:

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (1000 lbs)
[, 7] qsec 1/4 mile time
[, 8] vs V/S
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Looking at this list I would pick the following variables as having a potential influence on mpg; cyl, disp, hp, drat, wt. Let's build some models and compare them with anova before looking at tranmission types influence.

```
##Starting with weight as it is highly likely to have a strong influence an adding from there.
fit1 <- lm(mpg ~ wt, mtcars)
fit2 <- lm(mpg ~ wt + cyl, mtcars)
fit3 <- lm(mpg ~ wt + cyl + disp, mtcars)
fit4 <- lm(mpg ~ wt + cyl + disp + hp, mtcars)
fit5 <- lm(mpg ~ wt + cyl + disp + hp + drat, mtcars)

## Model comparison
anova(fit1, fit2, fit3, fit4, fit5)
```

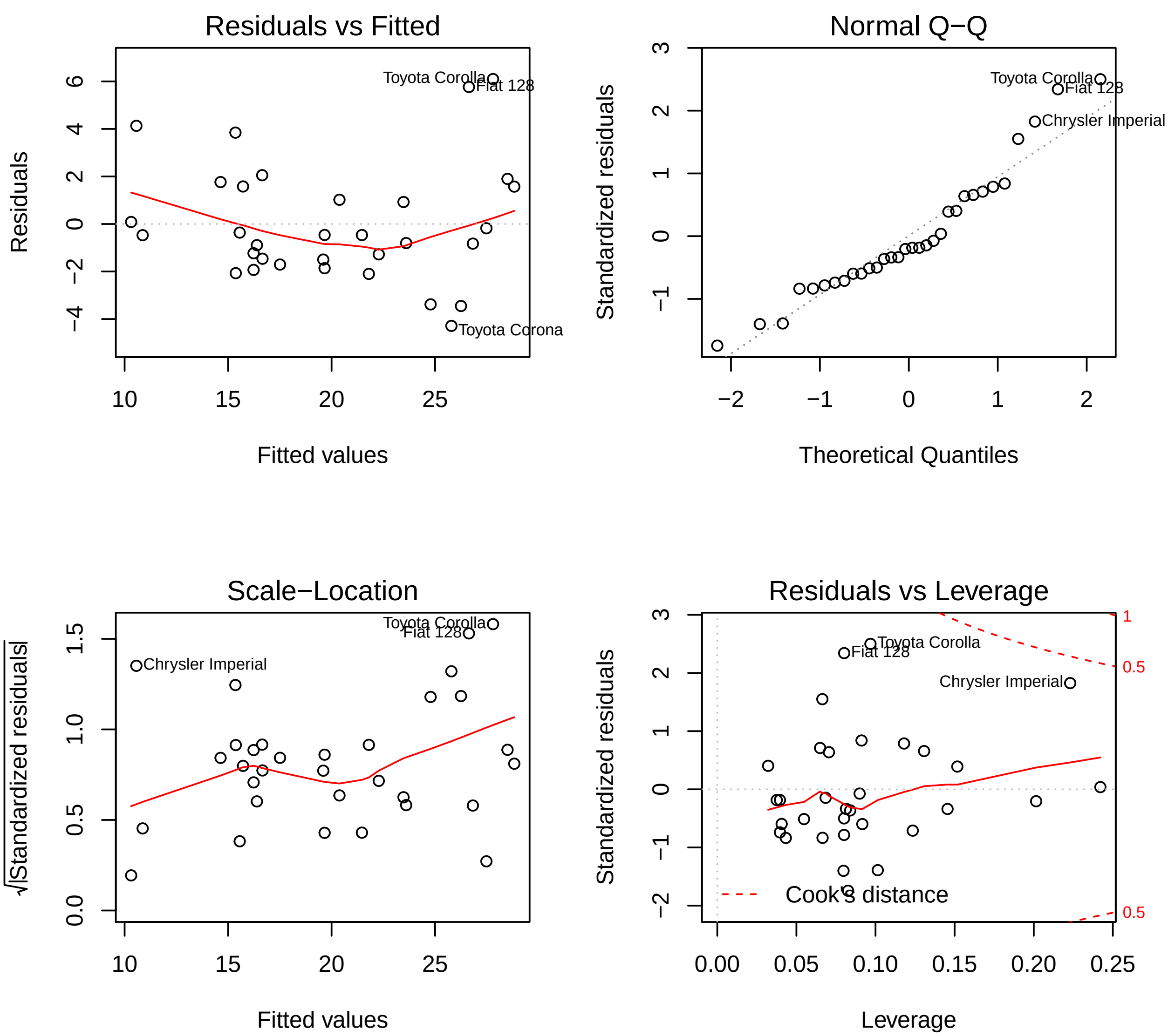
```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + cyl
## Model 3: mpg ~ wt + cyl + disp
## Model 4: mpg ~ wt + cyl + disp + hp
## Model 5: mpg ~ wt + cyl + disp + hp + drat
```

```
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 278.32
## 2     29 191.17  1    87.150 13.5337 0.001074 **
## 3     28 188.49  1     2.680  0.4161 0.524532
## 4     27 170.44  1    18.048  2.8027 0.106094
## 5     26 167.43  1     3.018  0.4687 0.499639
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I started with weight in the initial model and then added potential regressors 1 by 1. From a physics point of view weight has to influence mpg. Reviewing the models above it looks like adding cyl improved the model but all the additional variables after that didn't add any new insight. I would select model 2 as being the best fit from this group based on this.

Let's look at the residuals for this model and see if there is anything out of the ordinary.

```r
par(mfrow = c(2, 2))
plot(fit2)
```



So nothing here to indicate that our choosen model has any significant flaws.

Transmission type influence

Now let's get back to our original question, is an automatic or manual transmission better for MPG. Let's build on our second model by adding in the tranmission type and comparing it to the model without it.

```
fit_am <- lm(mpg ~ wt + cyl + am, mtcars)
summary(fit_am)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + am, 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 ***
## wt           -3.1251     0.9109  -3.431  0.00189 **
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## ammanual      0.1765     1.3045   0.135  0.89334
## ---
## 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
```

```
anova(fit2, fit_am)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + cyl
## Model 2: mpg ~ wt + cyl + am
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     29 191.17
## 2     28 191.05  1   0.12491 0.0183 0.8933
```

So comparing this new model with anova to our best fit without transmission type we see almost no change in RSS and a very low F score. Looking at our summary for the model with the tranmission included we can also see that the t value for tranmission is quite low. Even the Std. Error for it is significantly larger than the coefficient. So adding in tranmission type did almost nothing to improve our model

Conclusion

Looking at the data it appears that weight and the number of cylinders have the largest influence on MPG. Adding in tranmission type did nothing to improve our model. From this we can conclude that the tranmission type does not influence MPG one way or the other.