Analysis: Transmission choice and its impact on fuel efficiency

Stéphane Nyombayire

June 18, 2015

Executive Summary

The study aims to identify the impact of the design of transmission (automatic vs. manual) on gas consumption (MPG). Using the mtcars data set, we used two linear regression models (simple linear regression model and a multivariate regression model) we found that manual transmissions are indeed better for fuel economy than automatic transmissions. The multivariate model performed better than the singular based predictor model. This suggests that we could have enhanced our model to find the ultimate model that maximize fuel efficiency. This is however out of scope for this assignment.

Exploratory Analysis

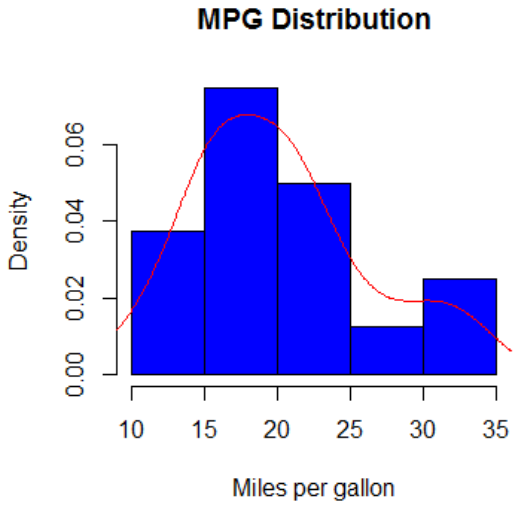
```
str(mtcars)
```

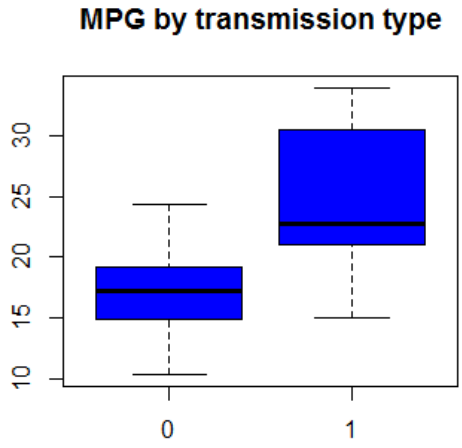
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 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 ...
```

Data frame with 32 observations on 11 variables. Variables of interests are mpg and am (0 = automatic, 1 = manual)

Histogram #1: MPG distribution shape with a fitted density over the image

Plot #2: Distribution of Manual versus Automatic transmission MPG





What is the most fuel efficient transmission?

1. Single predictor regression model

```
mtcars$am <- factor(mtcars$am)
fit <- lm(mpg ~ am , data = mtcars)
summary(fit)

##
## Call:
## lm(formula = mpg ~ 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 ***
## am1            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
```

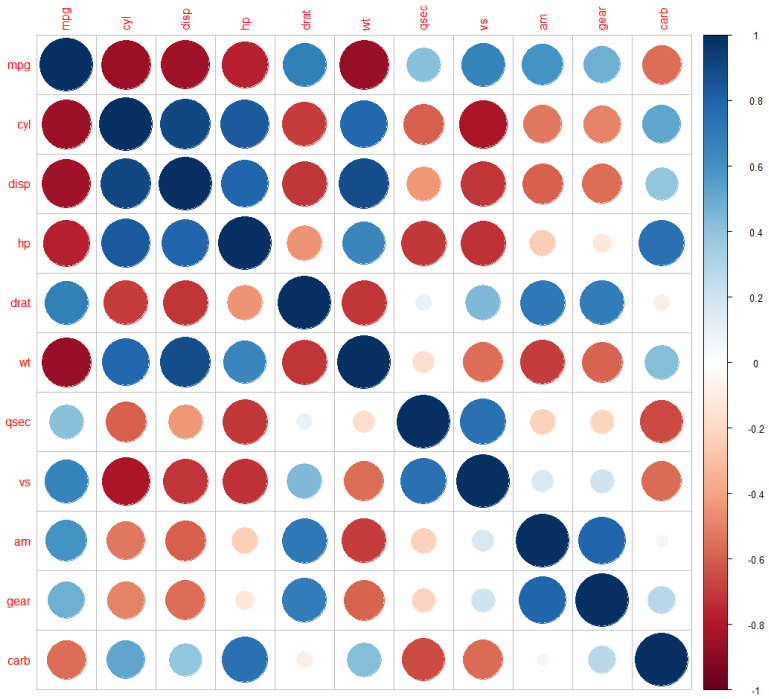
The P-Value is smaller than 0.05, which shows that transmission is a valid addition to the model. Examining the coeffiecients, we find that for this simple model, 7.244 is the increase in MPG between the automatic and manual transmission on average.Thus, Manual in this simplistic model improves fuel efficiency significantly over the automatic transmission.

2. Multivariate regression model

Let's now look at ways we could have improved our previous model and see whether Manual transmission is indeed the better choice when it comes to fuel efficiency. We will first find the important predictor and see which ones correlate with our response variable (i.e: MPG).

In order to visualize the correlation for the predictors, we will use corrplot to see how closely two pairs of predictors are correlated. As you can see below: cyl, hp, wt, disp are all negatively correlated to mpg.

```
corrplot(cor(mtcars))
```



Now, looking at the multivariable regression model with the additional predictors, we have the following:

```
mtcars$am <- factor(mtcars$am)
mtcars$cyl <- factor(mtcars$cyl)
fit2 <- lm(mpg ~ wt + am + cyl + hp + disp, data = mtcars)
sqrt(vif(fit2))
```

```
##          GVIF       Df GVIF^(1/(2*Df))
## wt   2.611892 1.000000        1.616135
## am   1.609627 1.000000        1.268711
## cyl  3.124944 1.414214        1.329568
## hp   2.176258 1.000000        1.475215
## disp 3.591864 1.000000        1.895221
```

```
fit2 <- lm(mpg ~ wt + am + cyl + hp, data = mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ wt + am + cyl + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## wt          -2.49683    0.88559  -2.819  0.00908 **
## am1          1.80921    1.39630   1.296  0.20646
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

For sanity check, we use vif to spot check any variance inflation amongst the factors. And as you can tell, both cyl and disp are pretty high which suggests that they are correlated. Thus, we removed disp in our final model. Examining the coeffiecients in the final model, we find that for this new model, 1.8 is the increase in MPG between the automatic and manual transmission on average while keeping all other predictors constant.

3. Comparing models

We will look at comparing the simple one predictor model to the multivariate one using ANOVA.

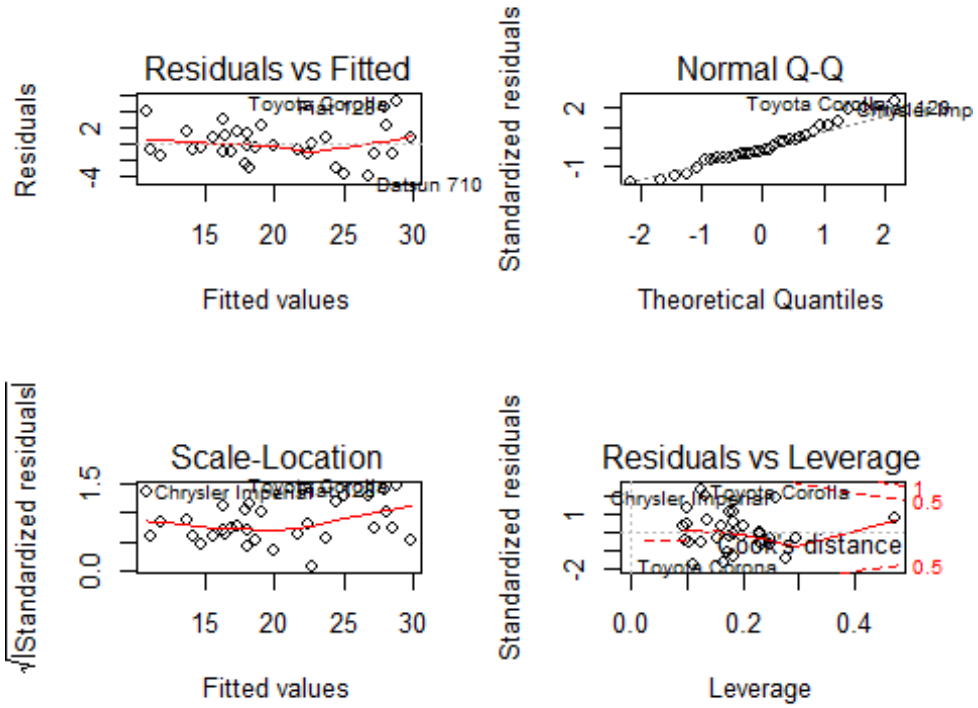
```
fit <- lm(mpg ~ am , data = mtcars)
fit2 <- update(fit, mpg ~  am + wt + cyl + hp)
anova(fit,fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + cyl + hp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By looking at the p Value is smaller than 0.01 so we know that the additional predictors are a welcome additions to our second model.

In addition, by closely examining the residual vs. fitted plot, we see that the residuals are uncorrelated with the ???t, independent and identically distributed, with a mean of 0.

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



Conclusion

All the other predictors added to the model do not change the initial conclusion that manual transmission is a more efficient transmission type from the MPG perspective, so the answer for the first question is clearly: Manual is better than Automatic. For the second question, the multivariate model shows that - 1.8 is the increase in MPG between the automatic and manual transmission on average while keeping all other predictors constant.