Motor Trend car tests - Relationship between type of transmission and fuel consumption

Fernando Flores

January 28th, 2016

Executive Summary

The current report aims to investigate the relationship between the type of transmission and fuel consumption based on data of 32 cars tested by Motor Trend US magazine for an issue published in 1974. The main target is to assess if automatic or manual transmission is better for MPG (miles per US gallon), and quantify the consumption difference for the mentioned transmission types.

The study concludes that manual transmission is better for MPG in average by 2.08 miles/gallon, but always considering not only the transmission type, but also engine horsepower and car weight, to explain this variability more accurately.

Exploratory Data Analysis

The basic features of the dataset are summarized in appendix A1. More details can be found in the documentation online

The source dataset has 11 columns of information from 32 cars. Some of the variables are needing reformat for the sake of the analysis. For the specifics of columns `vs` and `am` see this post in StackOverflow

```
mtcars$cyl <- as.factor(mtcars$cyl); mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb); mtcars$vs <- as.factor(mtcars$vs)
levels(mtcars$vs) <- c("V-engine", "Straight engine")
mtcars$am <- as.factor(mtcars$am)      ## Transmission
levels(mtcars$am) <- c("Automatic", "Manual")
```

The boxplot of fuel consumption by transmission type can be found in appendix A2. It gives helpful information about how well manual and automatic transmission are for MPG in the cars tested, but in order to answer the questions of interest we have to fit regression models.

Regression Models

Primary model

As a first step we fit a primary linear model with Miles/US gallon (variable `mpg`) as outcome and transmission type (variable `am`) as regressor.

```
fitPrimary <- lm(mpg ~ am, data = mtcars)
```

The coefficient summary table can be found in appendix A3. In the model, 17.15 is the empirical MPG mean for cars with automatic transmission. For the case of manual transmission, the value of 7.24 is an additional increase to the mentioned MPG mean. The p-value is 0.0002, being <0.05 means this model is statistically significant.

By looking at the R-squared value we can see that the transmission type explains only 36% of the variability in MPG. This value is not good enough for our analysis, so we will have to use nested models to select a better one.

Secondary Models

After researching general information about the car features described on the dataset, the variables related to fuel consumption in practice are:

- `hp`: Gross horsepower, a unit of measurement of engine power.
- `wt`: Weight (lb/1000).
- `qsec`: 1/4 mile time, which can be a good measure of acceleration by proxy.
- `gear`: Number of forward gears, which is related to transmission.
- `carb`: Number of carburetors, the device in charge of air-fuel mix for better consumption.

Besides the transmission type (`am`), this 5 variables are included one by one in nested models to look for a better explanation of the MPG variability.

```
fit2 <- update(fitPrimary, mpg ~ am + hp)
fit3 <- update(fitPrimary, mpg ~ am + hp + wt)
fit4 <- update(fitPrimary, mpg ~ am + hp + wt + qsec)
fit5 <- update(fitPrimary, mpg ~ am + hp + wt + qsec + gear)
fit6 <- update(fitPrimary, mpg ~ am + hp + wt + qsec + gear + carb)
```

After the analysis of variance (table in appendix A4), the p-values for models 2 and 3 are statistically significant. Both are also parsimonious so they are good candidates for the definitive model.

Model selection

Coefficients and variability explained

The full summary of coefficients can be found in appendix A5. Because this are multivariable models, the adjusted R-squared is a better metric. In model 2 the transmission type explains 77% of the variability in MPG, while model 3 explains 82% and it's still parsimonious, even more than the previous one.

Therefore, in this setting, model 3 is better to predict MPG by fitting transmission type, horsepower and weight as regressors.

Residuals and diagnostics

After looking at the residuals and some diagnostics metrics, we can conclude there is no potential problems with the selected model. Plots of this features can be found in appendix A6. The Residuals vs. Fitted shows the points are randomly distributed, no hint of a pattern, meaning the points are statistically independent. Q-Q shows normal distribution of the residuals. Regarding potencial influence and leverage points, plot of Residual vs. Leverage suggests there is no potential outliers, as confirmed by the tables in appendix A7.

Conclusions

- **Is an automatic or manual transmission better for MPG?** In general, manual transmission is better for MPG, but horsepower and weight are very important factors to explain this improvement.

- **Quantify the MPG difference between automatic and manual transmissions.** Holding constant the variables horsepower and weight, manual transmissions are better in average by 2.08 miles/gallon. This value is extracted from the summary table of model 3, in appendix A5.

Appendix

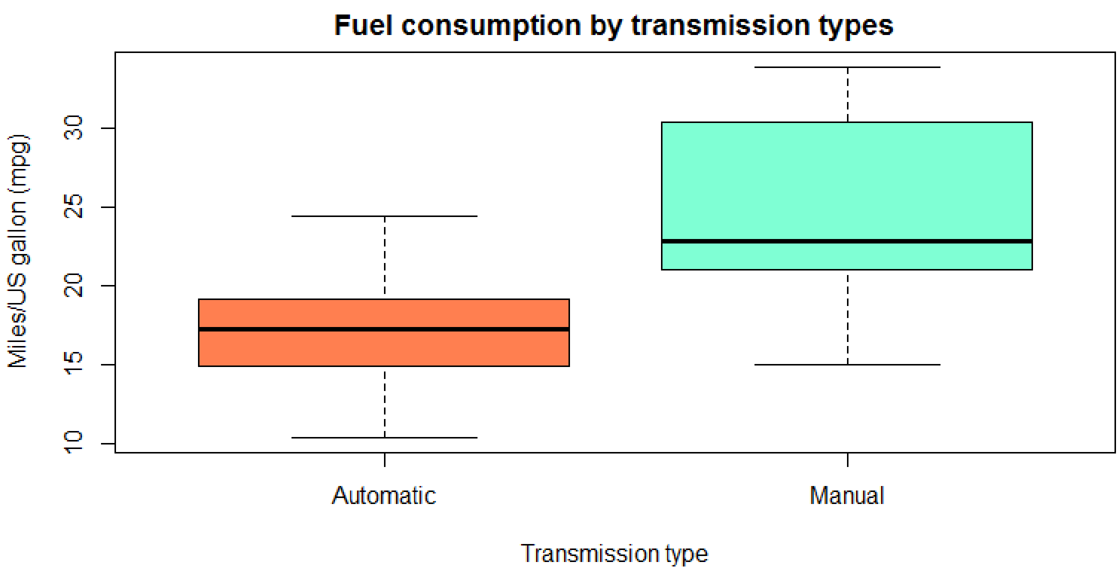
A1. Summary of the data

```
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 : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ 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  : Factor w/ 2 levels "V-engine","Straight engine": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

A2. Boxplot of fuel consumption by transmission type

```
par(mfrow = c(1, 1), mar = c(4, 4, 2, 1))
boxplot(mpg ~ am, data = mtcars, col = c("coral", "aquamarine"),
        xlab = "Transmission type", ylab = "Miles/US gallon (mpg)",
        main = "Fuel consumption by transmission types")
```



A3. Coefficients summary of the primary model

```
coeffPrimary

##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04

rSquaredPrimary

## [1] 0.3597989
```

A4. Summary of analysis of variance (ANOVA)

```
anova(fitPrimary, fit2, fit3, fit4, fit5, fit6)

## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + qsec
## Model 5: mpg ~ am + hp + wt + qsec + gear
## Model 6: mpg ~ am + hp + wt + qsec + gear + carb
```

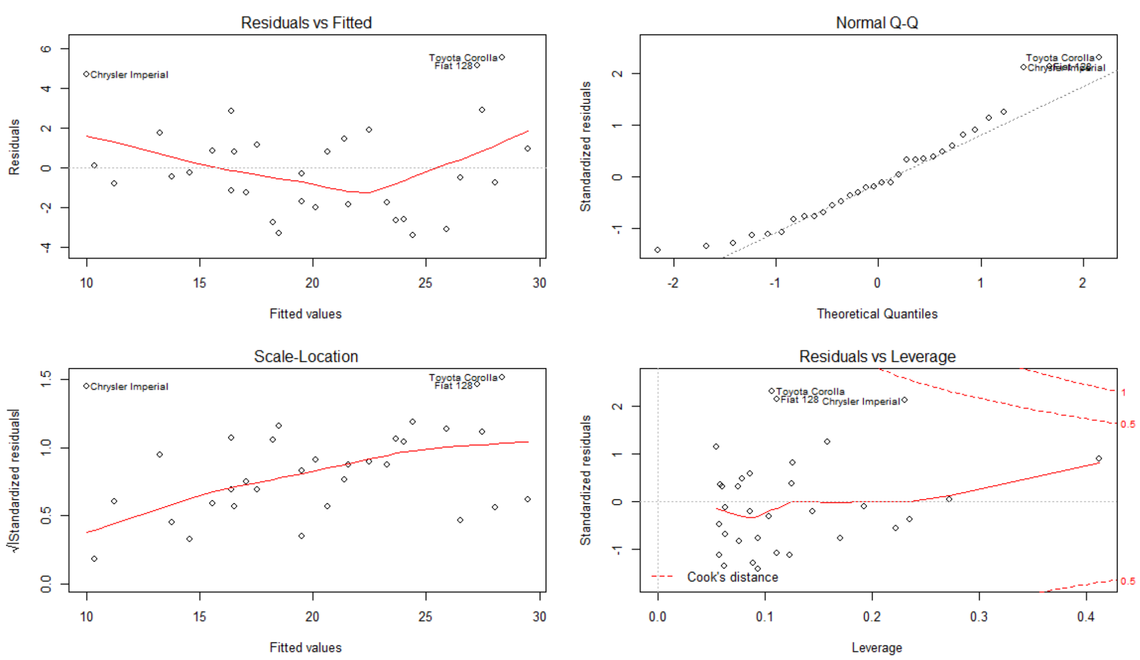
```
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)    
## 1     30 720.90                                   
## 2     29 245.44  1    475.46 62.9570 1.323e-07 ***
## 3     28 180.29  1     65.15  8.6265  0.008149 ** 
## 4     27 160.07  1     20.22  2.6780  0.117383    
## 5     25 158.30  2      1.76  0.1168  0.890394    
## 6     20 151.04  5      7.26  0.1923  0.961977    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A5. Coefficients summaries of models 2 and 3

```
summary(fit2)$coefficients

##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
## amManual     5.2770853 1.079540576  4.888270 3.460318e-05
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08

adjRSquared2

## [1] 0.7670025

summary(fit3)$coefficients

##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## amManual     2.08371013 1.376420152  1.513862 1.412682e-01
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03

adjRSquared3

## [1] 0.8227357
```

A6. Residuals and diagnostics plots

```
par(mfrow = c(2, 2), mar = c(5, 4, 2, 2))
plot(fit3)
```



A7. Summary of analysis for potential influence and leverage points

```
round(dfbetas(fit3)[, 2], 3)
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##              -0.335              -0.349              -0.230
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##              -0.054              -0.100               0.081
##          Duster 360           Merc 240D            Merc 230
##               0.032              -0.087              -0.098
##            Merc 280           Merc 280C          Merc 450SE
##               0.015               0.090              -0.017
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##              -0.044               0.060              -0.091
## Lincoln Continental   Chrysler Imperial            Fiat 128
##               0.009               0.540               0.432
##         Honda Civic      Toyota Corolla       Toyota Corona
##               0.019               0.223               0.286
##    Dodge Challenger         AMC Javelin          Camaro Z28
##               0.169               0.237               0.043
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##              -0.117              -0.035              -0.027
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##              -0.098              -0.086              -0.131
##       Maserati Bora          Volvo 142E
##               0.156              -0.306
```

All `dfbetas` are small, no points with high potential for influence.

```
round(hatvalues(fit3), 3)
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##               0.093               0.123               0.089
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##               0.075               0.079               0.076
##          Duster 360           Merc 240D            Merc 230
##               0.192               0.126               0.086
##            Merc 280           Merc 280C          Merc 450SE
##               0.063               0.063               0.059
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##               0.060               0.058               0.235
## Lincoln Continental   Chrysler Imperial            Fiat 128
##               0.273               0.230               0.111
##         Honda Civic      Toyota Corolla       Toyota Corona
##               0.125               0.107               0.170
##    Dodge Challenger         AMC Javelin          Camaro Z28
##               0.057               0.062               0.145
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##               0.054               0.104               0.086
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##               0.159               0.223               0.094
##       Maserati Bora          Volvo 142E
##               0.412               0.111
```

The point for `Maserati Bora` has high potential for leverage but no influence. Therefore, it can't be considered as outlier.