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.

## Exploratory Data Analysis

library(datasets)  
data(mtcars)

summary(mtcars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Min. :0.0000 Min. :3.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000   
## Median :0.0000 Median :4.000 Median :2.000   
## Mean :0.4062 Mean :3.688 Mean :2.812   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :5.000 Max. :8.000

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 ...

dim(mtcars)

## [1] 32 11

head(mtcars, 4)

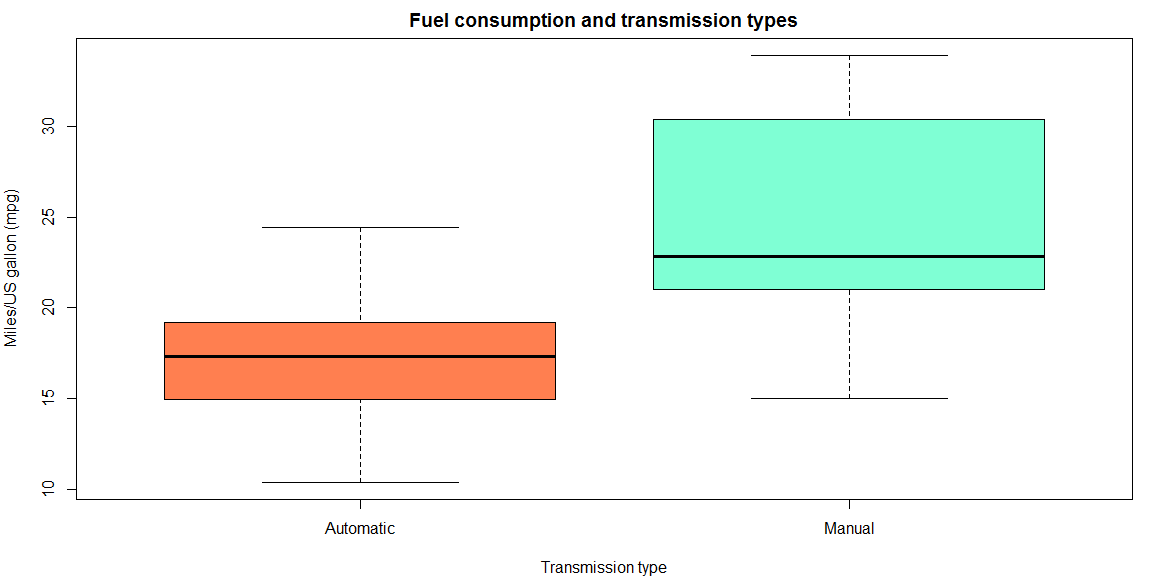
## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1

tail(mtcars, 4)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Ford Pantera L 15.8 8 351 264 4.22 3.17 14.5 0 1 5 4  
## Ferrari Dino 19.7 6 145 175 3.62 2.77 15.5 0 1 5 6  
## Maserati Bora 15.0 8 301 335 3.54 3.57 14.6 0 1 5 8  
## Volvo 142E 21.4 4 121 109 4.11 2.78 18.6 1 1 4 2

# Formatting factors for analysis (See help page ?mtcars)  
mtcars$cyl <- as.factor(mtcars$cyl)  
mtcars$gear <- as.factor(mtcars$gear)  
mtcars$carb <- as.factor(mtcars$carb)  
## See http://stackoverflow.com/questions/18617174/r-mtcars-dataset-meaning-of-vs-variable  
mtcars$vs <- as.factor(mtcars$vs)  
levels(mtcars$vs) <- c("V-engine", "Straight engine")  
mtcars$am <- as.factor(mtcars$am)  
levels(mtcars$am) <- c("Automatic", "Manual")

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 and transmission types")



*Figure N. Boxplot of MPG by transmission type*

## Regression Models

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

##   
## 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 \*\*\*  
## amManual 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

summary(fitPrimary)$coefficients

## 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

summary(fitPrimary)$adj.r.squared

## [1] 0.3384589

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)  
  
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

# Model 3 is parsimonious and statistically significant  
summary(fit3)$coefficients

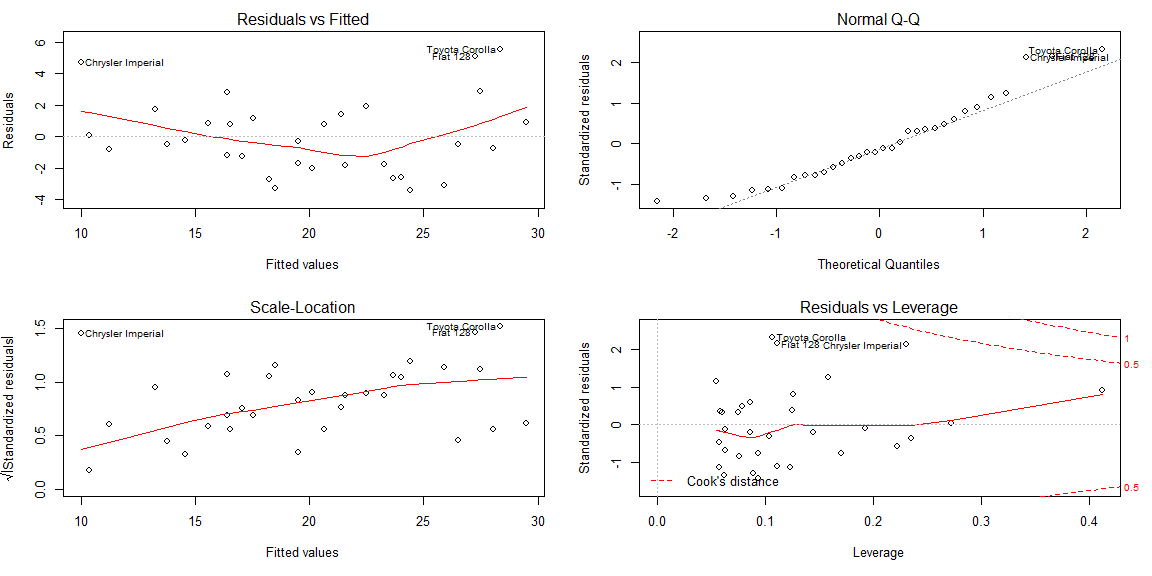
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13  
## amManual 2.08371013 1.376420152 1.513862 1.412682e-01  
## hp -0.03747873 0.009605422 -3.901830 5.464023e-04  
## wt -2.87857541 0.904970538 -3.180850 3.574031e-03

summary(fit3)$adj.r.squared

## [1] 0.8227357

## 0.8227357 => am, hp and wt explains 82% of the variability...

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



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

### 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

### "Maserati Bora" has high leverage but no influence

## Conclusions

## Appendix

### A1. abcabcabc