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

Fernando Flores

January 28th, 2016

## Execute 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)

The basic features of the dataset are summarized in appendix A1. More details can be found in [the documentation online](https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html)

The source dataset has 11 columns of information from 32 cars. The information contained in some of the variables is needing reformat for the sake of the current analysis. For the specifics of columns vs and am see [this post in StackOverflow](http://stackoverflow.com/questions/18617174/r-mtcars-dataset-meaning-of-vs-variable)

mtcars$cyl <- as.factor(mtcars$cyl) ## Number of cylinders  
mtcars$gear <- as.factor(mtcars$gear) ## Number of forward gears  
mtcars$carb <- as.factor(mtcars$carb) ## Number of carburetors  
mtcars$vs <- as.factor(mtcars$vs) ## Straight or V-engine  
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. This plot gives helpful information about how well manual and automatic transmission performs regarding MPG for the cars tested, but in order to answer the questions of interest we have to analyze 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 (variabla am) as regressor.

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

The coefficient summary table can be found in appendix A3. In the model, 17.1473684 is the empirical MPG mean for cars with automatic transmission. For the case of manual transmission, the value of 7.2449393 is an additional increase to the mentioned MPG mean.

According to the p-value, which is 310^{-4} and <0.05, 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 performance.

Besides the transmission type (am), this 5 variables are included one by one in nested models to look for a fitted linear model that can explain better 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

Because this are multivariable models, the adjusted R-squared is a better metric to evaluate variability.

adjRSquared2 <- summary(fit2)$adj.r.squared  
adjRSquared3 <- summary(fit3)$adj.r.squared

The full summary of coefficients can be found in appendix A5. In the second model, the transmission type explains 77% of the variability in MPG, while the third explains 82% and it's still parsimonious, even more than the previous one.

#### Residuals

#### Diagnostics

## Conclusions

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

## Appendix

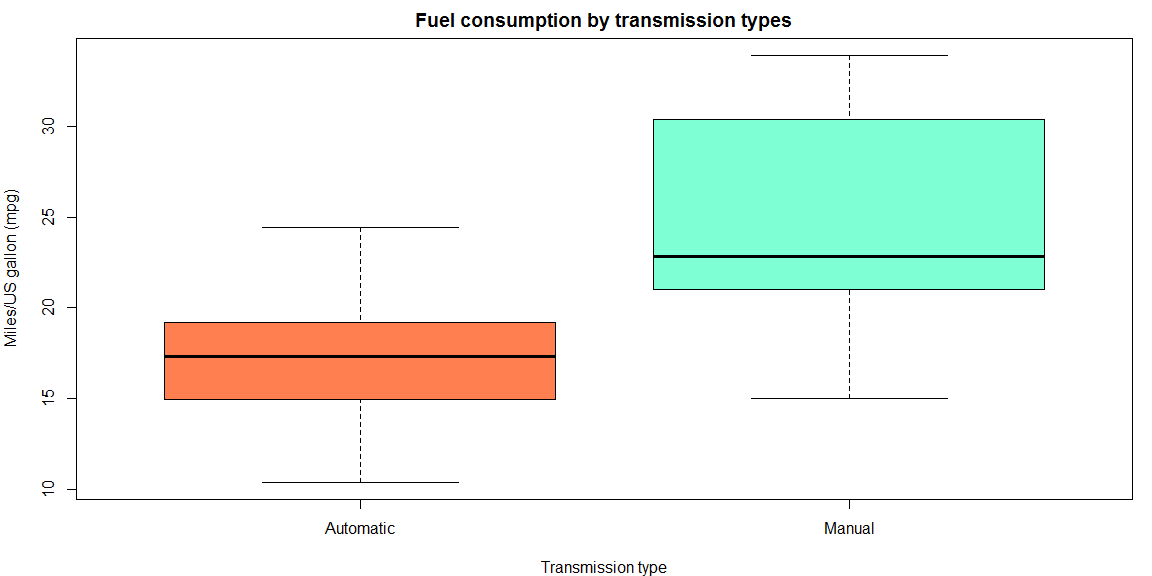
### A1. Summary of the data

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