

Study on the Relationship Between MPG and Transmission Type

Silva, RAFAEL

May 12, 2018

Setup

This section presents all the libraries used during the project.

```
library(ggplot2)
library(PerformanceAnalytics)
```

Executive Summary

This project looks into the *mtcars* dataset, with the objective of answering whether automatic or manual transmission is better for consumption in MPG and to quantify this difference.

Although in a direct comparison between automatic and manual transmission the manual transmission is better in 7.24 miles per gallon, when the influence of other correlated variables is considered, the data fails to reject the hypothesis of both transmissions having the same mpg with a P-Value of 0.88.

Introduction

The Motor Trend, a magazine about the automobile industry, is are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

1. “Is an automatic or manual transmission better for MPG”
2. “Quantify the MPG difference between automatic and manual transmissions”

The code chunk bellow loads the dataset into the work space. The dataset documentation is available [here](#) or by typing `?mtcars` into R console.

```
data("mtcars")
```

Exploratory Analysis

In order to better understand the data, a box plot was made showing how the mpg is distributed for each kind of transmission (Appendix A). The box plot makes it fairly obvious that there is a big difference in mpg on each kind of transmission.

Although helpful as an initial hunch, the box plot is not enough to take any conclusions on which kind of transmission is better regarding mpg. By isolating these 2 variables the results are possibly biased due to the omission of other correlated variables in the dataset.

The correlation chart (Appendix B) shows which of the other variables available in the dataset may introduce bias in the results of a direct comparison between both types of transmission. For instance, there is a significant correlation between the displacement and the transmission type, the negative correlation coefficient shows that the average displacement for the automatic cars in the dataset is bigger than for the manual cars. On the other hand, there is a negative correlation between the mpg and the displacement, which means that,

as displacement gets higher the mpg gets lower. This means that, at least part of the difference in mpg of both transmission types may be due to the manual cars of the dataset having a smaller engine.

Regression Models

The first thing to consider when comparing the effect of the transmission type is a direct comparison in mpg between both types, the results of this comparison is shown by the chunk bellow.

```
mtcars$amfac <- factor(mtcars$am, labels = c("automatic", "manual"))
lm0 <- lm(mpg ~ amfac, data = mtcars)
summary(lm0)$coef
```

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

The result of such comparison is that cars with manual transmission make on average 7.24 more miles per gallon than an automatic car as shown above.

However, as explained in the last section, this result may be biased by the influence of other variables. With the information gathered in the the exploratory analysis, an strategy was defined to select the best model for this comparison. Several nested models where fitted in the chunk bellow, one by one adding the variables that have a significant correlation with both mpg and am (minimum significance 0.05 = “*”), and then a final model with the rest of the variables where also fitted to show for the sake of testing the significance of the rest of the variables. An analysis of variance was made on all the models to which one is the best.

```
lm1 <- lm(mpg ~ amfac + gear, data = mtcars)
lm2 <- lm(mpg ~ amfac + gear + drat, data = mtcars)
lm3 <- lm(mpg ~ amfac + gear + drat + wt, data = mtcars)
lm4 <- lm(mpg ~ amfac + gear + drat + wt + disp, data = mtcars)
lm5 <- lm(mpg ~ amfac + gear + drat + wt + disp + cyl, data = mtcars)
lm6 <- lm(mpg ~ ., data = mtcars)
```

The results of the analysis of variance (Appendix C) on the models show that the only additions that where significant to the models were the rear axle ration (drat), the weight (wt), and the number of cylinders (cyl). The chunk bellow creates a final with the mentioned variables and prints its coefficients.

```
lmf <- lm(mpg ~ amfac + drat + wt + cyl, data = mtcars)
summary(lmf)$coef
```

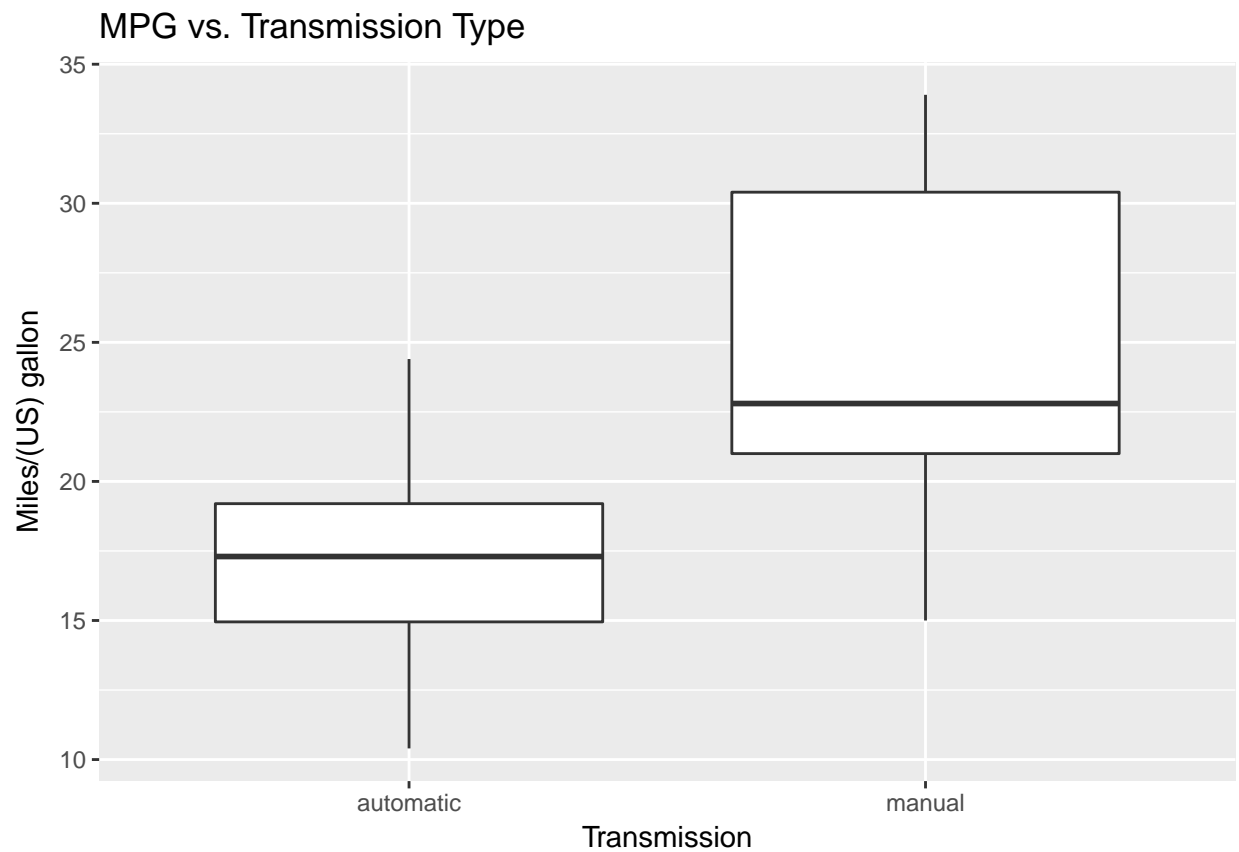
```
##              Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 39.9793439   7.1239820  5.61193777 5.916235e-06
## amfacmanual  0.2372691   1.5080241  0.15733776 8.761494e-01
## drat        -0.1301305   1.5290787 -0.08510387 9.328067e-01
## wt          -3.1326895   0.9317034 -3.36232479 2.322906e-03
## cyl         -1.5254047   0.4654059 -3.27757887 2.879278e-03
```

Conclusion

Based on the results of the regression model, it was concluded that although in a direct comparison between automatic and manual transmission the manual transmission is better in 7.24 miles per gallon, when the influence of other correlated variables is considered, the data fails to reject the hypothesis of both transmissions having the same mpg with a P-Value of 0.88.

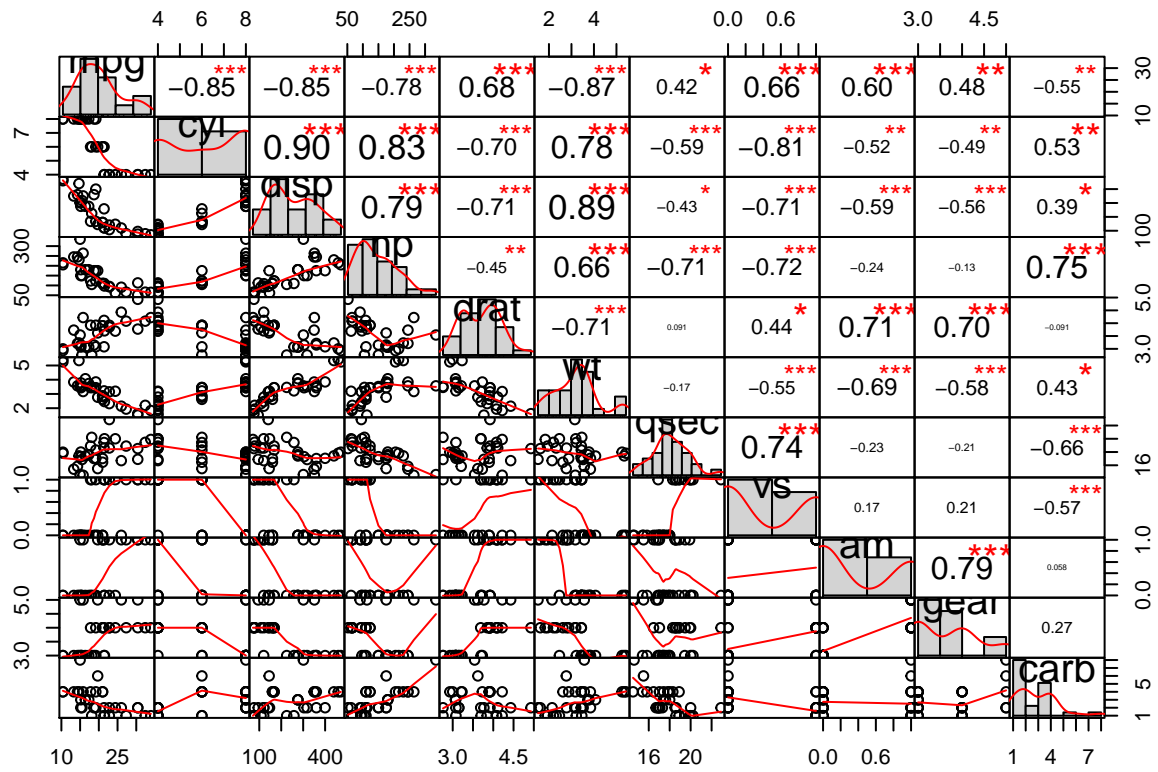
Appendix A

```
qplot(factor(mtcars$am, labels = c("automatic", "manual")), mpg, data = mtcars, geom = "boxplot",  
       xlab = "Transmission", ylab = "Miles/(US) gallon",  
       main = "MPG vs. Transmission Type")
```



Appendix B

```
chart.Correlation(mtcars[,1:11])
```



Appendix C

```
anova(lm0, lm1, lm2, lm3, lm4, lm5, lm6)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ amfac
## Model 2: mpg ~ amfac + gear
## Model 3: mpg ~ amfac + gear + drat
## Model 4: mpg ~ amfac + gear + drat + wt
## Model 5: mpg ~ amfac + gear + drat + wt + disp
## Model 6: mpg ~ amfac + gear + drat + wt + disp + cyl
## Model 7: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb +
##      amfac
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1      30 720.90
## 2      29 720.85  1      0.048  0.0069  0.93467
## 3      28 559.39  1    161.458 22.9881 9.749e-05 ***
## 4      27 260.41  1    298.984 42.5689 1.833e-06 ***
## 5      26 233.59  1     26.814  3.8177  0.06416 .
## 6      25 182.10  1     51.492  7.3313  0.01318 *
## 7      21 147.49  4     34.606  1.2318  0.32759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```