MPG consumption analysis betweeen automatic and manual transmissions

Ridzuan Mohamad

3/19/2017

Introduction

The Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they 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

Data Analysis & Exploratory

The dataset consist of 11 variables but we only interested only a few variable such as *mpg, cyl, hp, wt, am and gear*. The detail description for the variable as describe in **Appendex A and B**

```
mtcars_clean <- subset(mtcars, select = c("mpg", "cyl", "hp", "wt", "am", "gear"))
```

Than, calculate the mean for both transmission type. The boxplot graph as shown in **Appendix C**

```
c(automatic = mean(mtcars[mtcars$am==0,]$mpg),
    manual = mean(mtcars[mtcars$am==1,]$mpg))
```

```
## automatic    manual
##  17.14737  24.39231
```

The result shows that the automatic transmission is better than manual transmission in gas consumption. Let take that as our Null Hypothesis (H_0). Now we test the H_0 with T-Test.

```
t.test(mtcars[mtcars$am==0,]$mpg, mtcars[mtcars$am==1,]$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == 0, ]$mpg and mtcars[mtcars$am == 1, ]$mpg
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

The *p-value* from the T-Test is less than 0.005 so have to reject the H_0

Regression Model

Let creat 2 models for our case:-

1. model.simple which is mpg as outcome and am as predictor.
2. model.full which is mpg as outcome and other variables as predictor

```
model.simple <- lm(mpg ~ am, data = mtcars_clean)
model.all <- lm(mpg ~ ., data = mtcars_clean)
```

```
anova(model.simple, model.all)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am + gear
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 169.73  4    551.17 21.108 7.471e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary

Appendix

Appendix A

Variable	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
hp	Gross horsepower
wt	Weight (1000 lbs)
am	Transmission (0 = automatic, 1 = manual
gear	Number of forward gears

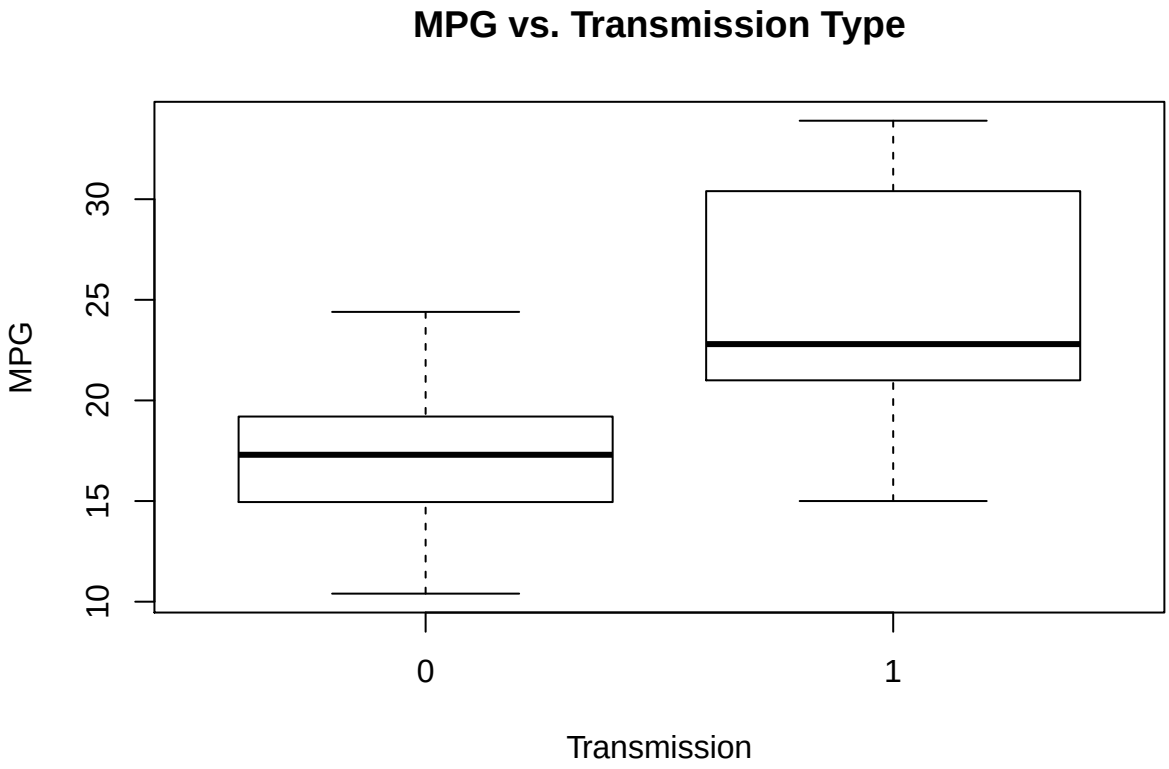
Appendix B

```
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 ...
```

Appendix C

```
with(mtcars, boxplot(mpg ~ am, xlab="Transmission", ylab="MPG",
        main="MPG vs. Transmission Type"))
```



Appendix D

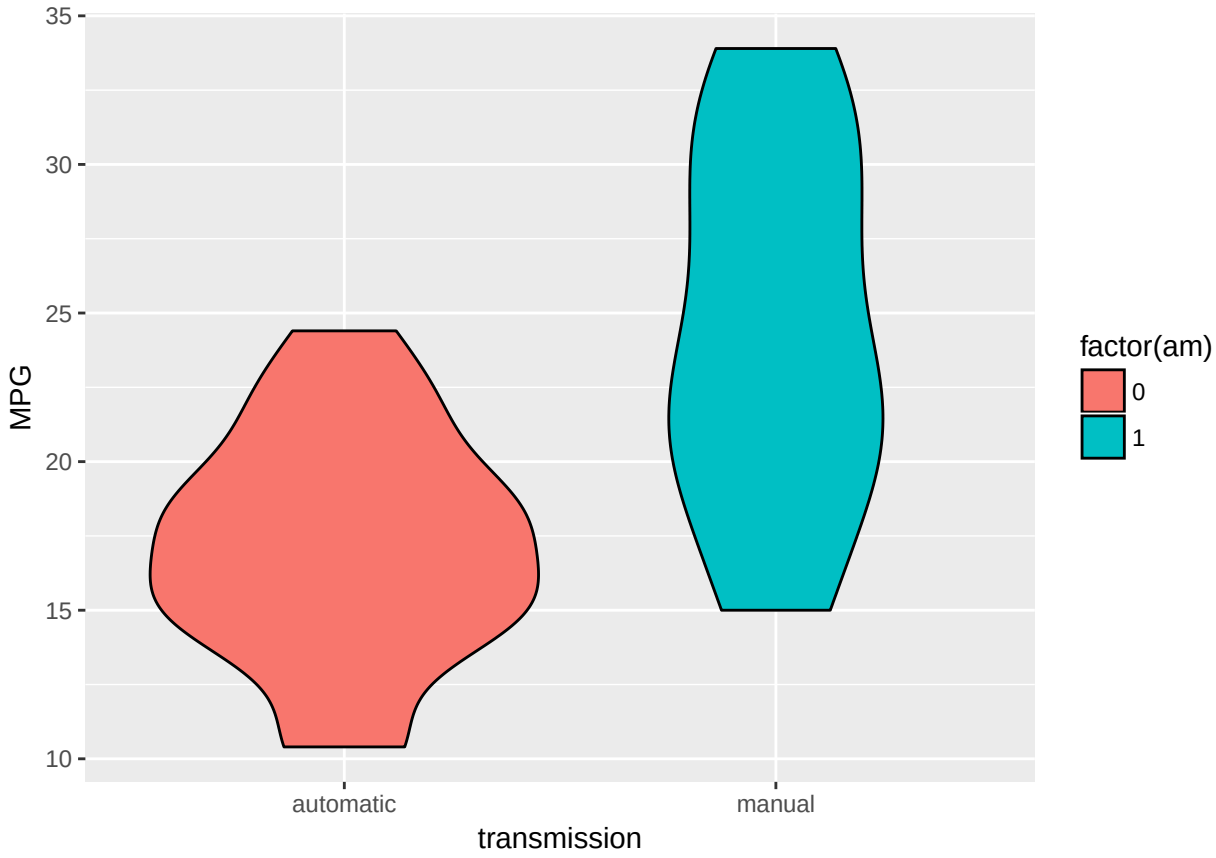
```
summary(model.simple)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars_clean)
##
## 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 ***
## am             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
```

```r
summary(model.all)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars_clean)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5843 -1.7328 -0.6248  1.2602  5.5846
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.18730    6.00091   6.197 1.48e-06 ***
## cyl         -0.80602    0.66416  -1.214  0.23581
## hp          -0.02339    0.01586  -1.475  0.15218
## wt          -2.63130    0.94448  -2.786  0.00983 **
## am           1.66574    1.73190   0.962  0.34501
## gear        -0.24116    1.18188  -0.204  0.83991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.555 on 26 degrees of freedom
## Multiple R-squared:  0.8493, Adjusted R-squared:  0.8203
## F-statistic:  29.3 on 5 and 26 DF,  p-value: 6.687e-10
```

```r
ggplot(mtcars_clean, aes(y=mpg, x=factor(am, labels = c("automatic", "manual")), fill=factor(am)))+
        geom_violin(colour="black")+
        xlab("transmission") + ylab("MPG")
```


35
30
25
20
15
10
MPG
automatic
manual
transmission
factor(am)
0
1