MPG consumption analysis betweeen automatic and manual transmissions

Ridzuan Mohamad

3/19/2017

Introduction

The Motor Trend, a magazine about the automobile industry are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). The data set of collection of cars (**mtcars**) is used for this study, The focus 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 and 32 observation data. The detail description for each variable as per *Appendix A*. Now, let look on the dataset structure.

```
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 ...
```

Than, calculate the mean for both transmission type and plot the boxplot to visualized the pattern for both transmission types. *(refer Appendix B)*

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

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

The result from mean calculation and boxplot shows that the automatic transmission is better than manual transmission in *mpg* consumption. Let take that as our Null Hypothesis (H_0). Now we test the H_0 with T-Test and the probability of Type 1 error is ($\alpha = 0.05$)

```
hypo.result <- t.test(mtcars[mtcars$am==0,]$mpg, mtcars[mtcars$am==1,]$mpg)
hypo.result$p.value
```

```
## [1] 0.001373638
```

However the hypothesis testing show that the *p-value* $< \alpha$ which force us to reject the H_0.

Regression Model

Firstly, let look at the linear model that corespond to our H_0.*(Appendix D)*

```
# factor variable am
mtcars$am <- as.factor(mtcars$am)

#create a simple model
model.0 <- lm(mpg ~ am, mtcars)
result.0 <- summary(model.0)
```

The Adjusted R-squared = 0.3384589. It shows that the model only able to explain 34% of the variance.

So, we have to include other variable to our model by using R step() function.

```
best.model <- step(lm(mpg ~ ., data = mtcars), trace=0)
result.1 <- summary(best.model)
result.1
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am1           2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The Adjusted R-squared = 0.8335561 shows that 83% of the variance of the MPG variable. Which mean the variable *wt, qsec and am* have a correlation affect on *mpg*.

Summary

Based on the Coefficients table for *best.model* and looking at categorical variables ***am1***, we can assumed that the manual transmission is 2.9358372 mpg better than automatic transmission.

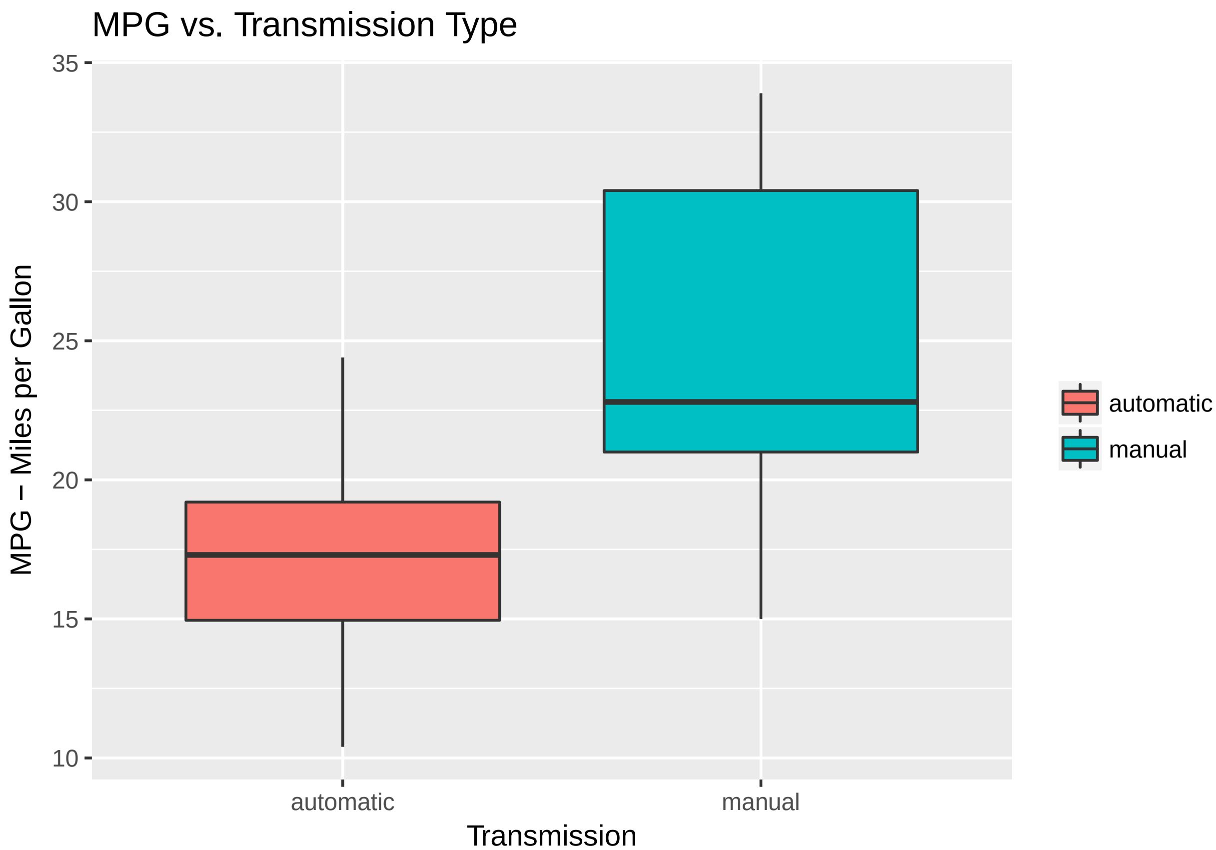
Appendix

Appendix A - Selected Variable and Description

No.	Variable	Description
1.	mpg	Miles/(US) gallon
2.	cyl	Number of cylinders
3.	disp	Displacement (cu.in.)
4.	hp	Gross horsepower
5.	drat	Rear axle ratio
6.	wt	Weight (1000 lbs)
7.	qsec	1/4 mile time
8.	vs	V/S
9.	am	Transmission (0 = automatic, 1 = manual
10.	gear	Number of forward gears
11.	carb	Number of carburetors

Appendix B - Boxplot MPG vs. Transmission Type

```
ggplot(mtcars, aes(x = factor(am, labels=c("automatic", "manual")), y = mpg,
                   fill=factor(am, labels=c("automatic", "manual")))) +
    geom_boxplot() +
    scale_x_discrete(name = "Transmission") +
    scale_y_continuous(name = "MPG - Miles per Gallon") +
    ggtitle("MPG vs. Transmission Type") +
    theme(legend.title=element_blank())
```



Appendix D - Model's Summary

```
summary(lm(mpg ~ am, mtcars))
```

```
##
## 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 ***
## am1            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(best.model)
```

```
##
```

```
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am1           2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Appendix E - Residuals

```r
par(mfrow = c(2,2))
plot(best.model)
```