

# Interaction Between Transmission Type and MPG

KTezel

12/19/2020

## Executive Summary

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

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

From the dataset, it is clear that there is a significant difference between the average mpg per automatic and manual transmission. To show the differences between average mpg per transmission type, a boxplot method is used to show the differences between median mpgs. Then, hypothesis testing, specifically, T-test is used to see whether the difference between the differences in means are significant or not. According to the T-test, differences were significant and that the mpg for automatic cars were higher than that of the cars with manual transmission.

## Libraries

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
```

## Getting the Data

```
data(mtcars)
```

## Data Preprocessing

```
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

## Exploratory Data Analysis

### Looking at data briefly

```
head(mtcars)
```

```
##           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
## Hornet Sportabout 18.7    8  360 175  3.15  3.440 17.02  0  0    3    2
## Valiant        18.1    6  225 105  2.76  3.460 20.22  1  0    3    1
```

```
dim(mtcars)
```

```
## [1] 32 11
```

### Looking for missing data

There is no missing data.

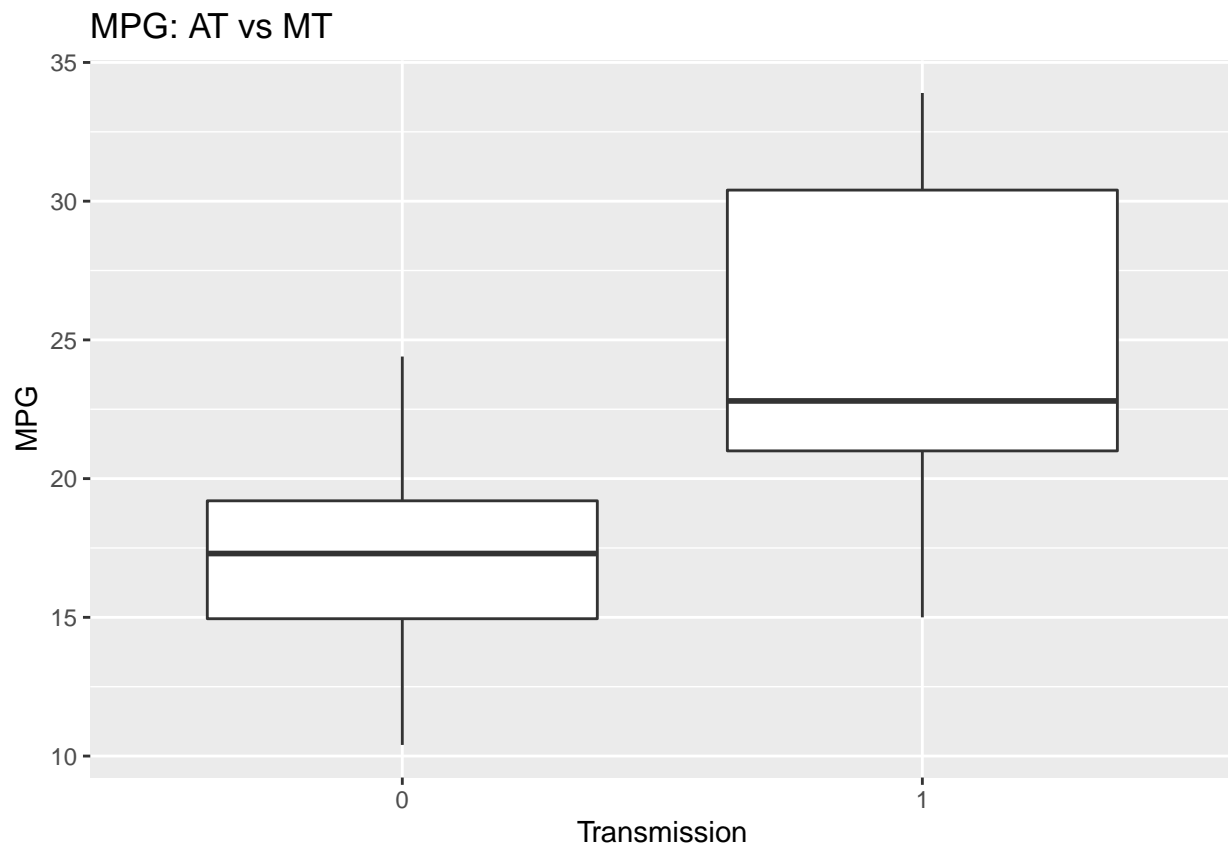
```
colSums(is.na(mtcars))
```

```
##  mpg  cyl  disp  hp  drat    wt  qsec  vs  am  gear  carb
##    0    0    0    0    0    0    0    0    0    0    0
```

### Plotting MPG and Automatic Transmission (1 = Yes, 2 = No)

From the plot graph, it looks like when the transmission is automatic, MPG increases. Below is the boxplot graph. It can be clearly seen that the median MPG is higher for the cars with automatic transmission.

```
p2 <- ggplot(data = mtcars, aes(factor(am),mpg)) + geom_boxplot() + labs(x= "Transmission", y = "MPG", title = "MPG: AT vs MT")
p2
```



## Inference

Are the differences in mean mpg between automatic and manual is significant?  
Let's use T-test to verify.

```
t.test(mpg ~ am, data = mtcars)
```

```
##
## Welch Two Sample t-test
##
## data: mpg by am
## 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 in group 0 mean in group 1
## 17.14737 24.39231
```

P-value from the t-test is very small which suggest that our null hypothesis is false, hence suggesting that the difference in mean mpg between automatic and manual is pretty significant.

## Regression Analysis

### Checking to see the relationship between MPG and Transmission

#### Linear Regression

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

##
## 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
```

Linear regression model predicts that when the car is automatic, it increases the mpg by approximately 7 miles. Intercept and the coefficients are significant and the variables explain approximately 36% (R-squared 0.3598) of the variation in MPG which is not a good model since 35% R-squared value is not good enough.

### Other Regression Models

Below are some other models that has some higher R-squared ratio. For the purposes of this report, only the R-squared values will be extracted from the report.

```
summary(lm(mpg~wt*factor(cyl),data=mtcars))$r.squared
```

```
## [1] 0.861561
```

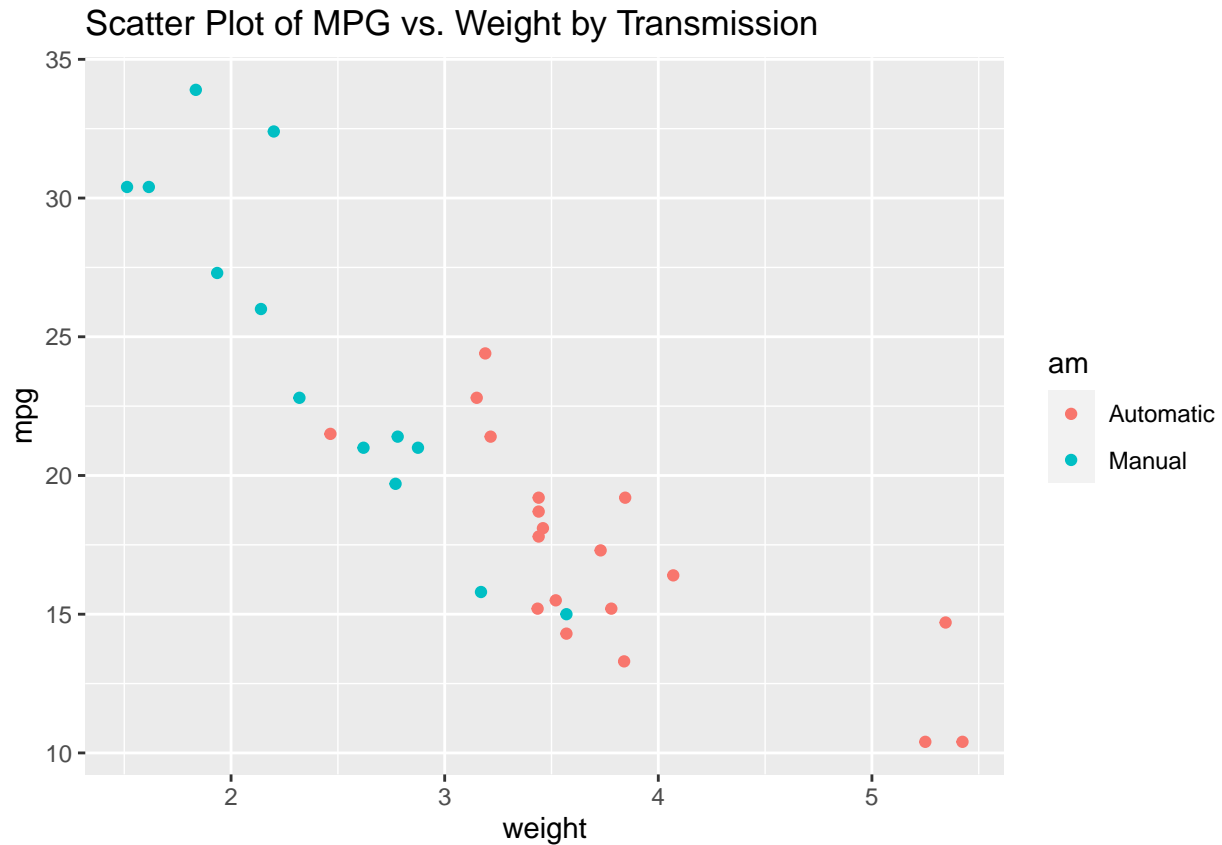
```
summary(lm(mpg~wt+factor(cyl),data=mtcars))$r.squared
```

```
## [1] 0.8374325
```

## Appendix

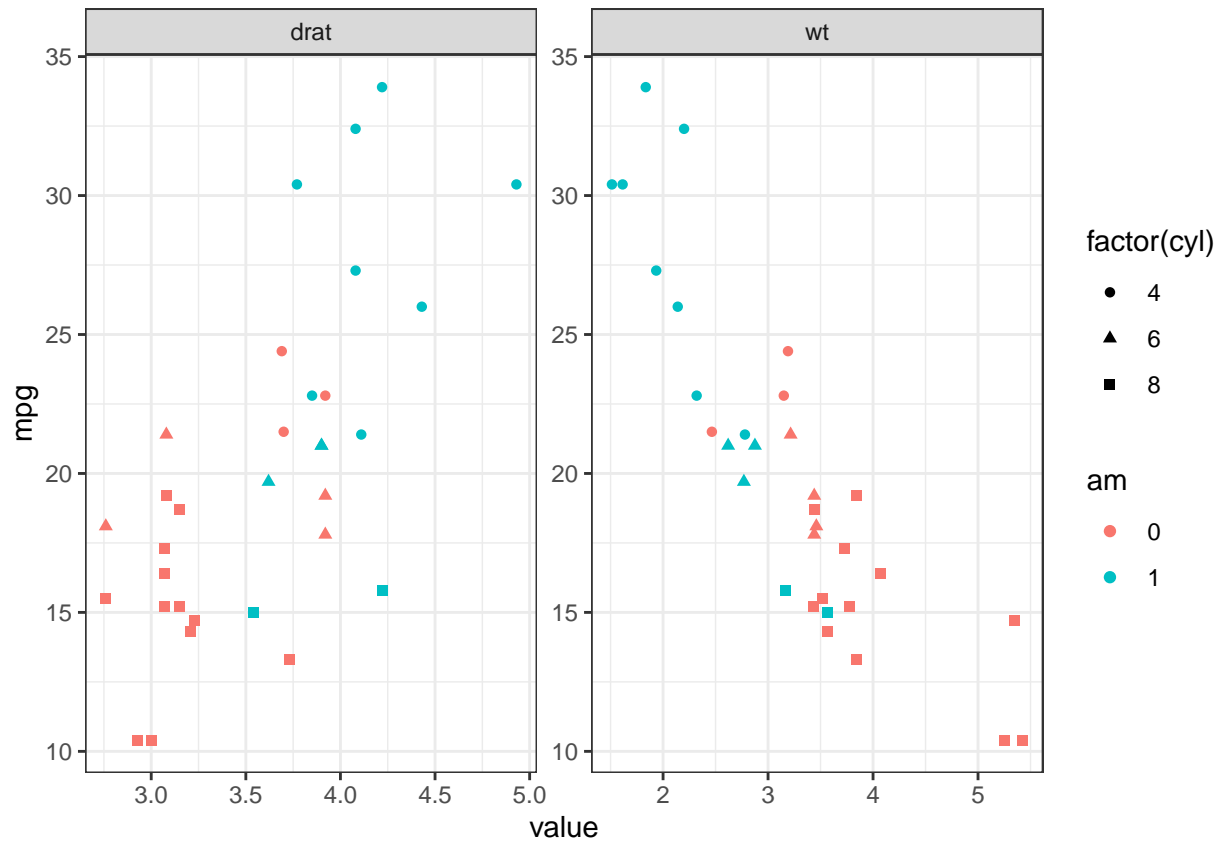
```
ggplot(mtcars, aes(x=wt, cyl, drat, y=mpg, group=am, color=am, height=3, width=3)) + geom_point() +  
scale_colour_discrete(labels=c("Automatic", "Manual")) +  
xlab("weight") + ggtitle("Scatter Plot of MPG vs. Weight by Transmission")
```

## Warning: Duplicated aesthetics after name standardisation:



```
mtcars_subset <- subset(mtcars, select = c(mpg, cyl, wt, drat, am))
```

```
mtcars_subset %>%  
  gather(-mpg, -am, -cyl, key="var", value = "value") %>%  
  ggplot(aes(x = value, y = mpg, color = am, shape = factor(cyl))) +  
  geom_point() +  
  facet_wrap(~ var, scales = "free") +  
  theme_bw()
```



## Residual Plot for the best model

```
fit_1 <- lm(mpg~wt+factor(cyl),data=mtcars)
residPlot <- ggplot(aes(x=.fitted, y=.resid), data=fit_1) + geom_point() + geom_hline(yintercept=0)+labs(
residPlot
```

