Automatic vs Manual: Which Transmission Has Better MPG?

Regression Models Course Project

Harkishan Grewal

12/05/2020

Executive Summary

This project uses the *mtcars* data set to examine the relationship between miles-per-gallon (MPG) and other variables. The project attempts to determine which transmission (automatic or manual) is better for MPG, and quantifies the difference.

The Regression Models section draws inference with a simple linear regression model and a multiple regression model. Both models show that manual transmission cars, on average, have a significantly higher MPG than automatic transmission cars.

In the linear model, the mean difference is 7.245 MPG; the average MPG for automatic transmission cars is 17.147 MPG, and the average MPG for manual transmission cars is 24.392 MPG. In the multiple regression model, the MPG difference is 2.9358 MPG.

The linear model shows that transmission type only accounts for 36% of the MPG performance. The multivariable regression model the other variables explain 85% of the MPG performance.

Exploratory Analysis and Visualizations

```
# Load Datasets
library(datasets)
data(mtcars)

# Get overview of dataset
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
```

```
# Get overview of Variables
str(mtcars)
```

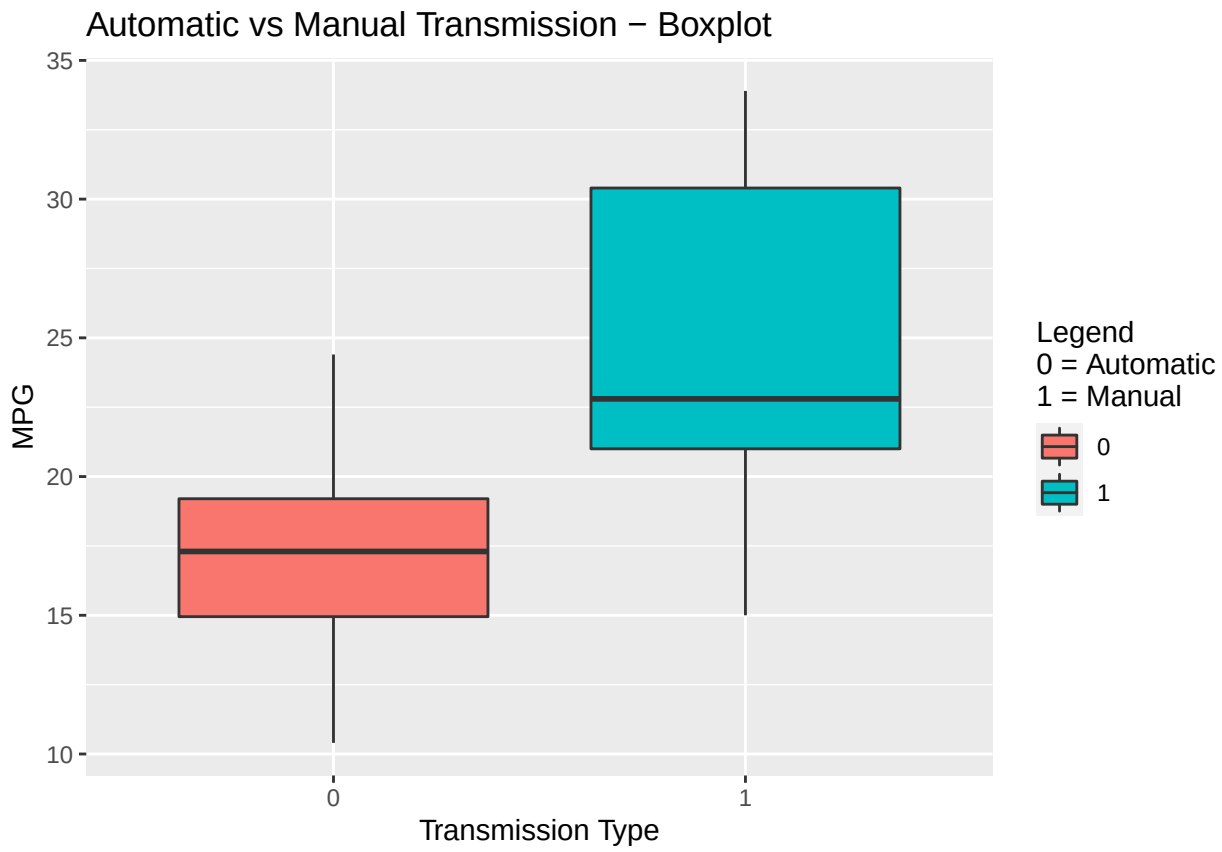
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 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 ...
```

```
# Statistical Summary of MPG variable
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.43   19.20   20.09   22.80   33.90
```

```
# Automatic vs Manual Transmission - Boxplot
library(ggplot2)
mtcars$am <- as.factor(mtcars$am)
trans_type <- ggplot(aes(x=am, y=mpg), data=mtcars) + geom_boxplot(aes(fill=am))
trans_type <- trans_type + labs(title = "Automatic vs Manual Transmission - Boxplot")
trans_type <- trans_type + xlab("Transmission Type")
trans_type <- trans_type + ylab("MPG")
trans_type <- trans_type + labs(fill = "Legend \n0 = Automatic\n1 = Manual")
trans_type
```




```
# Automatic vs Manual Transmission - Hypothesis Test:
auto <- mtcars[mtcars$am == "0",]
man <- mtcars[mtcars$am == "1",]
t.test(auto$mpg, man$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  auto$mpg and man$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
```

Regression Models

Linear Regression

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

```
##
## 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
```

Multivariable Regression

```
mrModel <- lm(mpg~am + cyl + hp + wt, data = mtcars)
summary(mrModel)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4765 -1.8471 -0.5544  1.2758  5.6608
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 36.14654    3.10478  11.642 4.94e-12 ***
```

```
## am1          1.47805    1.44115   1.026   0.3142
## cyl         -0.74516    0.58279  -1.279   0.2119
## hp          -0.02495    0.01365  -1.828   0.0786 .
## wt          -2.60648    0.91984  -2.834   0.0086 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.509 on 27 degrees of freedom
## Multiple R-squared:  0.849,  Adjusted R-squared:  0.8267
## F-statistic: 37.96 on 4 and 27 DF,  p-value: 1.025e-10
```

Linear vs Multivariable Regression

```
anova(lrModel, mrModel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + hp + wt
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.9
## 2     27 170.0  3     550.9 29.166 1.274e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

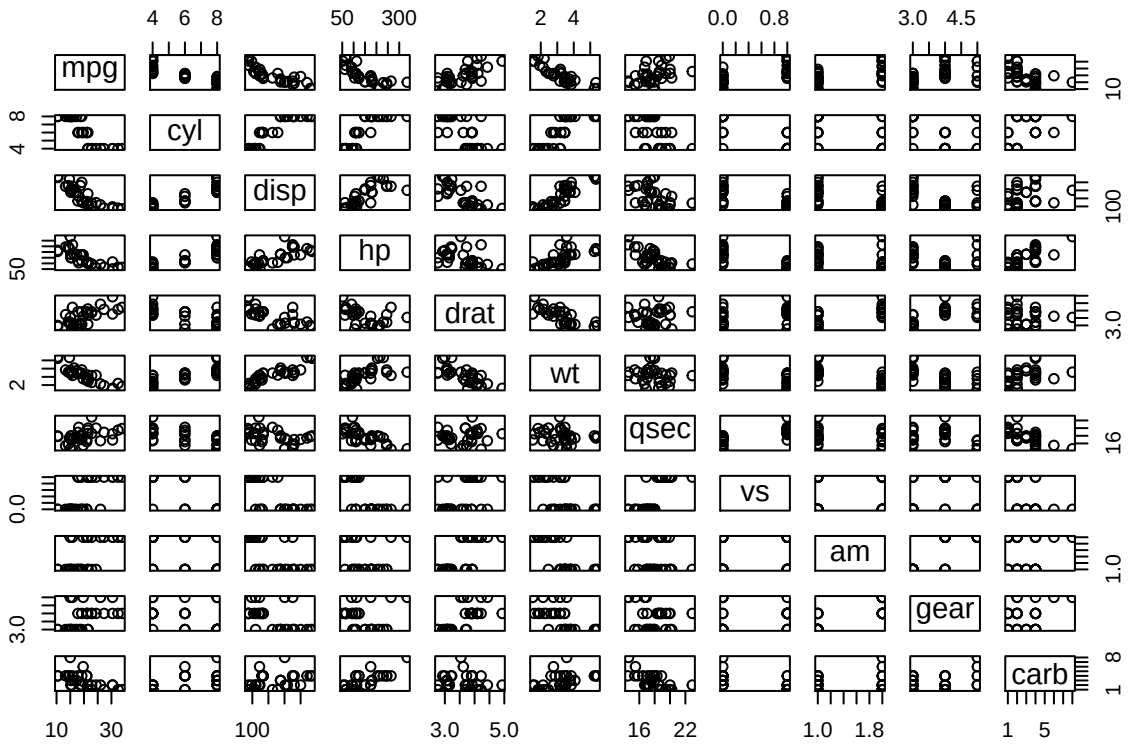
Conclusion

A linear regression model is used to determine the effect of transmission type on gas mileage (MPG). The model shows that manual transmission cars provide 7.25MPG (the am1 coefficient) better performance than automatic cars. Based on the R-squared value, transmission type only accounts for 36% of the MPG performance. Thus, a linear regression model is not sufficient to answer Motor Trend's questions with high certainty.

A multivariable regression model is used to account for the effect other variables on gas mileage (e.g. number of cylinders, engine horsepower, vehicle weight, etc.) This model showed a 1.48MPG increase from manual transmission cars over automatic cars with the additional variables explaining 85% of the MPG performance.

Appendix

```
# Scatterplot matrix of the mtcars dataset
pairs(mpg ~ ., data = mtcars)
```




```
# Scatterplots of the multivariable regression model residuals
par(mfrow = c(2,2))
#plot(mrModel)
```