

Motor Trend Magazine MPG Study

Jeffrey Strickland

1/13/2022

Executive Summary.

This report answers questions pertaining to Motor Trend, a magazine about the automobile industry. They wanted to understand the relationship between a set of variables and miles per gallon (MPG) (outcome). They were 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 results show that manual transmission vehicles provide better gas mileage than automatic transmissions. It also shows quantifies the MPG difference between automatic and manual transmissions, with acceleration and vehicle weight as additional factors affecting gas mileage.

Data Processing

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

```
library(datasets)
data(mtcars)
help(mtcars)
```

Data Description

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

It consists of 32 observations on 11 variables.

1. [1] mpg Miles/(US) gallon
2. [2] cyl Number of cylinders
3. [3] disp Displacement (cu.in.)
4. [4] hp Gross horsepower
5. [5] drat Rear axle ratio
6. [6] wt Weight (lb/1000)
7. [7] qsec 1/4 mile time
8. [8] vs Engine (0 = V-shaped, 1 = straight)
9. [9] am Transmission (0 = automatic, 1 = manual)
10. [10] gear Number of forward gears
11. [11] carb Number of carburetors

Exploratory Data Analysis (EDA)

The data summaries for automatic and manual transmissions are listed in Appendix A.

```
library(dplyr)
my_quantile <- function(x, probs) {
  tibble(x = quantile(x, probs), probs = probs)
}
mtcars %>%
  group_by(am) %>%
  summarise(xbar = mean(mpg), sd = sd(mpg))

## # A tibble: 2 x 3
##       am  xbar  sd
##   <dbl> <dbl> <dbl>
## 1     0  17.1  3.83
## 2     1  24.4  6.17
```

More

```
mtcars %>%
  group_by(am) %>%
  summarise(my_quantile(mpg, c(0.25, 0.5, 0.75)))

## # A tibble: 6 x 3
## # Groups:   am [2]
##       am      x probs
##   <dbl> <dbl> <dbl>
## 1     0  15.0  0.25
## 2     0  17.3  0.5
## 3     0  19.2  0.75
## 4     1  21    0.25
## 5     1  22.8  0.5
## 6     1  30.4  0.75
```

Data Preparation

To perform this analysis, we need to transform the class of some variables. The function `factor()` is used to encode a vector as a factor.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

1. Is an automatic or manual transmission better for MPG?

Here we quantify the MPG difference between automatic and manual transmissions. First, we generate a boxplot of MPG by transmission types at Appendix B. The boxplot shows that vehicles with manual transmissions may achieve better gas mileage than vehicles with automatic transmissions.

Hypothesis Test

Now we address Question 1 statistically. The null hypothesis we are testing is:

Ho: There is no different in mpg between automatic and manual transmissions.

We conduct a t-test to test this hypothesis.

```
t_test<-t.test(mtcars$mpg~mtcars$am, conf.level = 0.95)
df <- data.frame(
  "Welch Two Sample t-test"=rep(c('t-test p-value', 'Lower CI', 'Upper CI', 'Automatic
Estimate', 'Manual Estimate',''), times=1),
  Value=rep(c(sprintf("%.6f", t_test$p.value), sprintf("%.6f", t_test$conf.int),
    sprintf("%.6f", t_test$estimate), times=1)))
knitr::kable(df)
```

Welch.Two.Sample.t.test	Value
t-test p-value	0.001374
Lower CI	-11.280194
Upper CI	-3.209684
Automatic Estimate	17.147368
Manual Estimate	24.392308

Based on the results, $p\text{-value} = 0.001374 < 0.05$, we reject the null hypothesis that there is no difference in MPG, and conclude that manual transmission is better than automatic transmission for MPG, with assumption that all other conditions remain unchanged.

2. Quantify the MPG difference between automatic and manual transmissions

Here we try to quantify the MPG difference between transmission types, and find if there are other variables that account for the MPG differences.

First, do a multivariate linear regression with all variables. We use the step function in R for a step-wise regression, where the choice of predictor is carried out automatically by comparing certain criterion, for example, Akaike information criterion (AIC).

```
stepmodel = step(lm(data = mtcars, mpg ~ .), trace=0, steps=10000)
summary(stepmodel)

##
## 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 ***
## amManual      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
```

To further optimize the model, we can examine $\text{mpg} \sim \text{wt} + \text{qsec}$ correlation with am .

```

model <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model)

##
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.9692     5.7756   2.419  0.02259 *
## factor(am)Automatic:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)Manual:wt      -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)Automatic:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)Manual:qsec     1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13

```

The results suggests that the best model includes weight and qsec. The Adjusted R-squared increased from 0.83 to 0.88. The adjusted R-squared increases when the new term improves the model more than would be expected by chance. Thus, the model improved by adding the factors. that the model explains 88% of the variability of the response data around its mean. Now we can make the following conclusions:

- when the weight increased by 1000 lbs, the mpg decreased by -3.176 for automatic transmission cars, and -6.09 for manual transmission cars
- with increasing car weight we should choose automatic transmission cars for better gas mileage
- when the acceleration speed dropped, and 1/4 mile time increased (by 1 sec), the mpg factor increased by 0.834 miles for automatic transmission cars, and 1.446 miles for manual transmission cars
- with lower acceleration speed and same weight, manual transmission cars get better gas mileage mpg

Residual plots seems to be randomly scattered, and some transformation may be needed for linearity (Appendix C). The Normal Q-Q plot shows the residuals are approximately normally distributed. Both scale-location and residual vs leverage show no issues.

Conclusion

The hypothesis tests in part 1 showed that manual transmission vehicles are better than automatic transmission vehicles with no other factors considered. The factors in the optimized model quantifies the difference between automatic and manual transmissions, based on vehicle weight and acceleration, shows that manual transmission vehicles provide better gas mileage (increased mpg).

Appendix A

Data summary for automatic transmissions:

```
summary(mtcars[mtcars$am==0,])

##      mpg      cyl      disp      hp      drat
##  Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA
## 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
##  Median : NA   Median : NA   Median : NA   Median : NA   Median : NA
##  Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN
## 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
##  Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA
##      wt      qsec      vs      am      gear
##  Min.   : NA   Min.   : NA   Min.   : NA   Automatic:0   Min.   : NA
## 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   Manual    :0   1st Qu.: NA
##  Median : NA   Median : NA   Median : NA                      Median : NA
##  Mean   :NaN   Mean   :NaN   Mean   :NaN                      Mean   :NaN
## 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA                      3rd Qu.: NA
##  Max.   : NA   Max.   : NA   Max.   : NA                      Max.   : NA
##      carb
##  Min.   : NA
## 1st Qu.: NA
##  Median : NA
##  Mean   :NaN
## 3rd Qu.: NA
##  Max.   : NA
```

Data summary for manual transmissions:

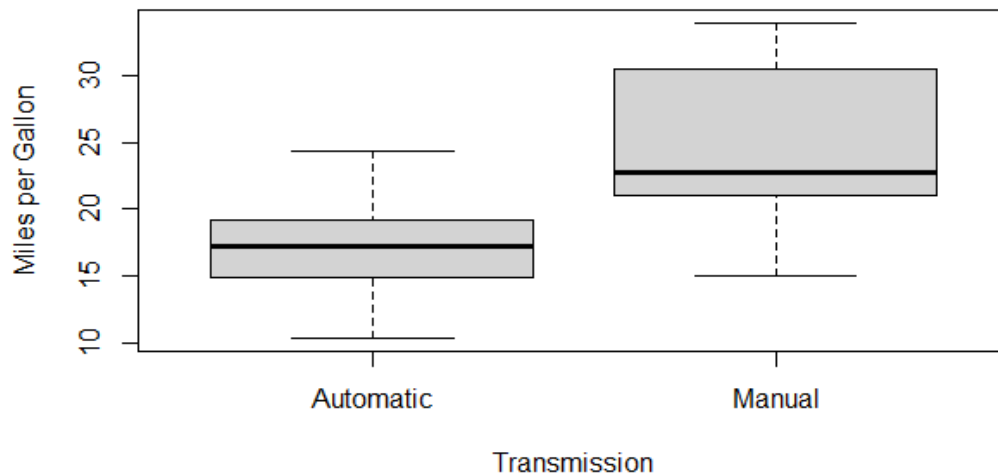
```
summary(mtcars[mtcars$am==1,])

##      mpg      cyl      disp      hp      drat
##  Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA   Min.   : NA
## 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
##  Median : NA   Median : NA   Median : NA   Median : NA   Median : NA
##  Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN   Mean   :NaN
## 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
##  Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA   Max.   : NA
##      wt      qsec      vs      am      gear
##  Min.   : NA   Min.   : NA   Min.   : NA   Automatic:0   Min.   : NA
## 1st Qu.: NA   1st Qu.: NA   1st Qu.: NA   Manual    :0   1st Qu.: NA
##  Median : NA   Median : NA   Median : NA                      Median : NA
##  Mean   :NaN   Mean   :NaN   Mean   :NaN                      Mean   :NaN
## 3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA                      3rd Qu.: NA
##  Max.   : NA   Max.   : NA   Max.   : NA                      Max.   : NA
##      carb
##  Min.   : NA
## 1st Qu.: NA
##  Median : NA
##  Mean   :NaN
## 3rd Qu.: NA
##  Max.   : NA
```

Appendix B

```
boxplot(mpg~am, data = mtcars, xlab = "Transmission", ylab = "Miles per Gallon",
main = "MPG by Transmission Type")
```

MPG by Transmission Type



Appendix C

Residual plots

```
par(mfrow = c(2,2))  
plot(model)
```

