

Motor Trend Analysis

Parker Oakes

June 4, 2019

Executive Summary

The purpose of the following analysis is to statistically address two issues for *Motor Trend* magazine:

1. Is an automatic or manual transmission better for MPG
2. Quantify the MPG difference between automatic and manual transmissions

The data provided in this analysis is the “mtcars” dataset that “was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 11 aspects of automobile design and performance for 32 automobiles (1973–74 models)” ¹.

The format of the dataset’s variables are as follows:

[, 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 Engine (0 = V-shaped, 1 = straight)
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Exploratory Data Analysis

Since the R Documentation (<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html>) states that the variable “am” represents automatic transmission and manual transmission (in respective order), I assigned these qualitative levels to these two categories. To determine a good baseline of difference between the transmissions, I chose to find the average of MPG per transmission. When subtracting the automatic from the manual, the difference came to manual transmission has 7.2449 more MPG on average than the automatic given 32 vehicles.

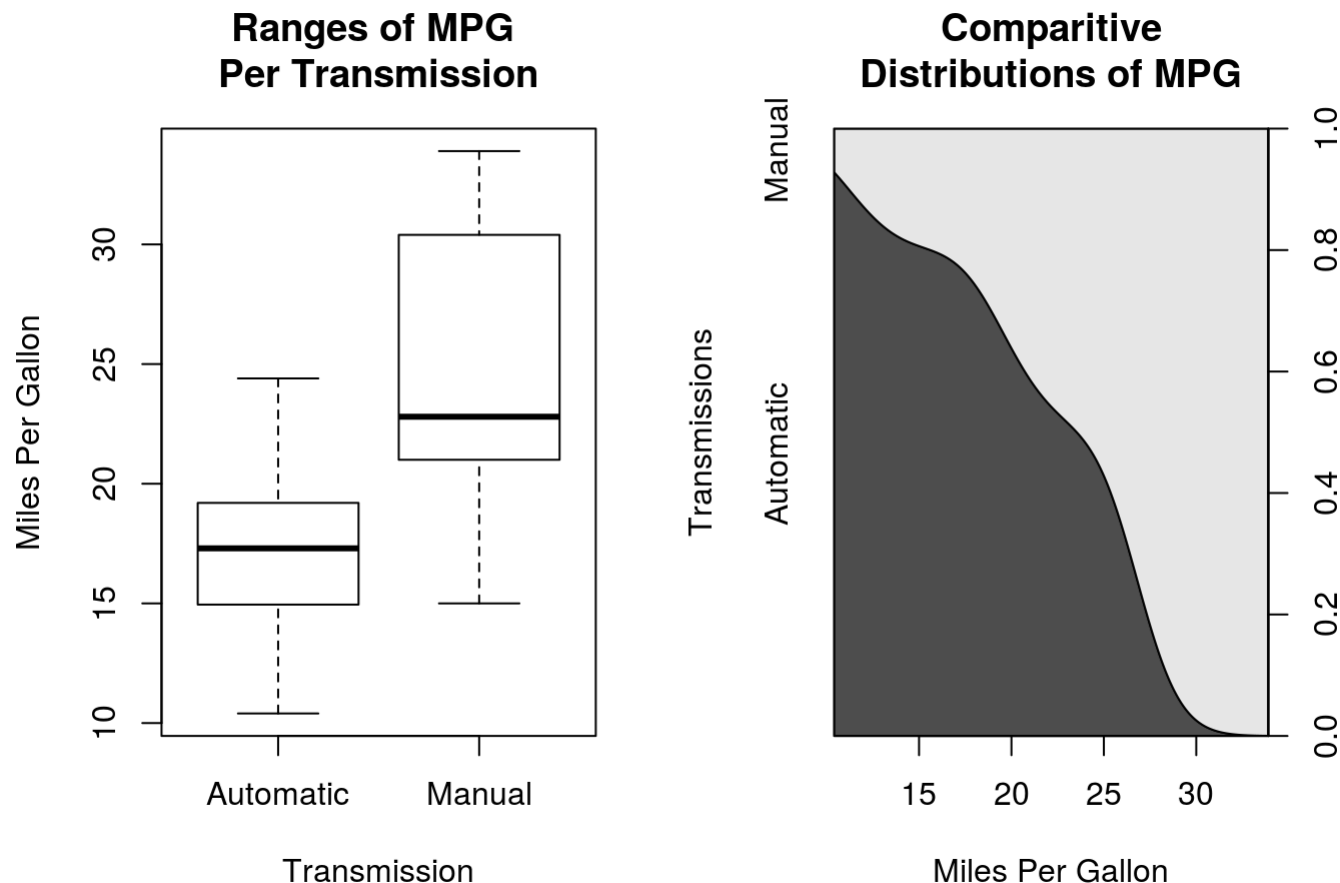
```
data(mtcars)
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
with(mtcars, tapply(mpg, am, mean))
```

```
## Automatic    Manual
##  17.14737    24.39231
```

```

par(mfrow = c(1,2))
with(mtcars,boxplot(mpg ~ am,
                    names = c("Automatic","Manual"),
                    xlab = "Transmission",
                    ylab = "Miles Per Gallon",
                    main = "Ranges of MPG \nPer Transmission"))
cdplot(as.factor(mtcars$am) ~ mtcars$mpg,
       main = "Comparitive\nDistributions of MPG",
       xlab = "Miles Per Gallon",
       ylab = "Transmissions")

```



Per the above (EDA) Exploratory Data Analysis plots that determine the averages of MPG per automatic and manual transmissions, it is conclusive that the manual transmission ranks on the higher end only by frequency. We are not able to ultimately conclude the answer to question 1 given frequency alone but must form a hypothesis given the explored data.

Hypothesis

$H_0 : MPG_{Manual} \leq MPG_{Automatic}$

$H_1 : MPG_{Manual} > MPG_{Automatic}$

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic","Manual")
Automatic <-mtcars[mtcars$am == "Automatic",]
Manual     <-mtcars[mtcars$am == "Manual",]
NullHypothesis <- t.test(Manual$mpg, Automatic$mpg)
NullHypothesis$p.value
```

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

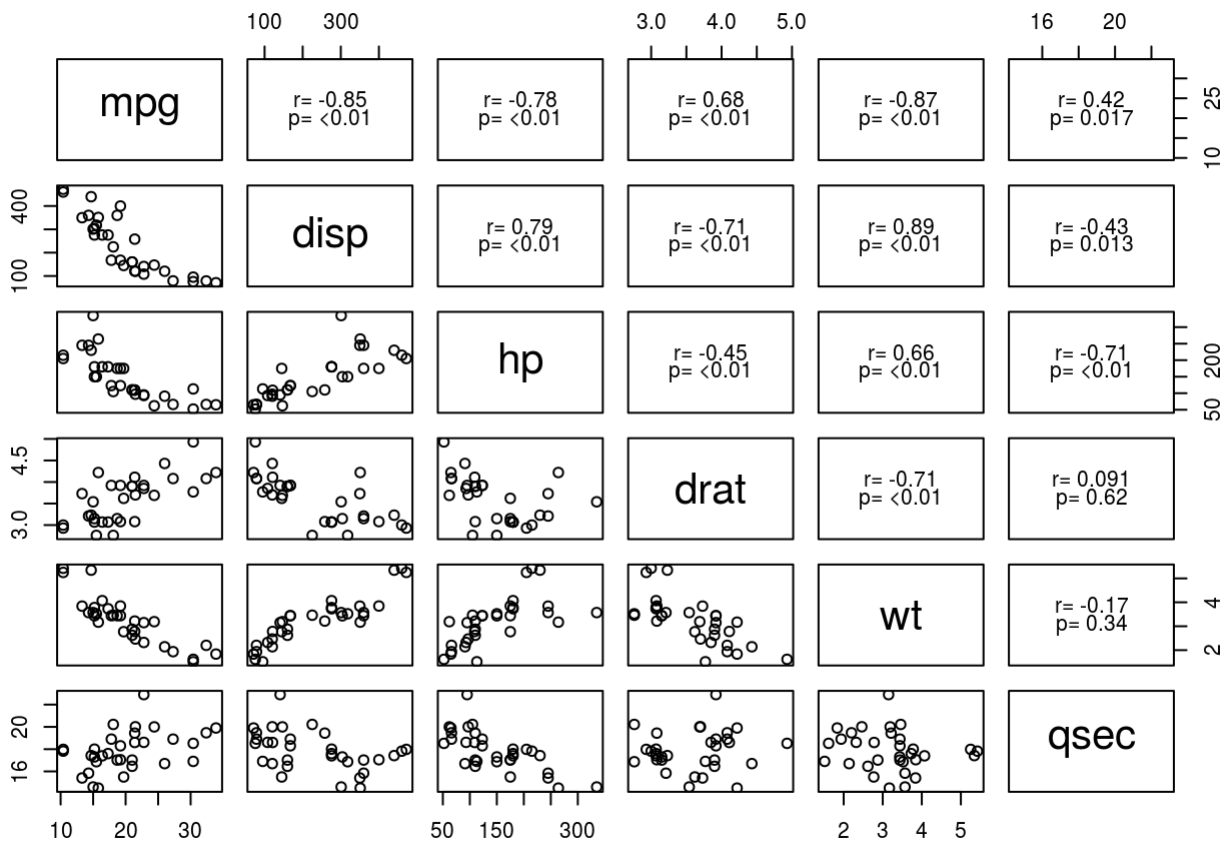
To answer question 1, it is concluded that manual transmissions are much better MPG. Our H_0 was rejected since the above p-value is 0.0014 and is 0.0486 less than 0.05.

Correlation Analysis

Next, to answer question 2 and quantify our results, we must determine correlations in our data. The “panel.cor” function was used in the upper panel of the “pairs” function found in base R².

```
mtcars1.1 <- mtcars[, c(1,3,4,5,6,7)]
panel.cor <- function(x, y, digits = 2, cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  # correlation coefficient
  r <- cor(x, y)
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste("r= ", txt, sep = "")
  text(0.5, 0.6, txt)

  # p-value calculation
  p <- cor.test(x, y)$p.value
  txt2 <- format(c(p, 0.123456789), digits = digits)[1]
  txt2 <- paste("p= ", txt2, sep = "")
  if(p<0.01) txt2 <- paste("p= ", "<0.01", sep = "")
  text(0.5, 0.4, txt2)
}
pairs(mtcars1.1, upper.panel = panel.cor)
```



As we can tell by the above correlation plot, most variables that have the strongest correlation to MPG are negatively sloped (e.g. displacement in cu. in., horse power, weight). I chose column set [1, 3, 4, 5, 6, 7] since this variable set is related to the car and the column set [2, 8, 10, 11] is only related to number of moving parts within the engine.

Regression Modeling

Model 1: Single Linear Regression

```
Model_1 <- lm(mpg ~ am, data = mtcars)
summary(Model_1)$r.squared; summary(Model_1)$adj.r.squared
```

```
## [1] 0.3597989
```

```
## [1] 0.3384589
```

From this Single Variable Regression model of mpg and am, the R^2 value of this model is 0.3598, meaning that it only explains 35.98% of the variance of the data.

Model 2: Multivariate Linear Regression

```
Model_2 <- step( lm (data = mtcars, mpg ~ .), trace=0, steps=10000)
summary(Model_2)$r.squared; summary(Model_2)$adj.r.squared
```

```
## [1] 0.8496636
```

```
## [1] 0.8335561
```

The above second model is using the step function with the linear model function to determine the variables that best affect mpg of the 32 cars. Per the above multivariate model, The R^2 of the model is 0.8497 meaning it is explaining 84.97% of the data.

Model 3: ANOVA of Models 1 & 2

```
Model_2.1 <- lm( mpg ~ am + wt + qsec, data = mtcars)
Model_3 <- anova(Model_1, Model_2.1)
format(Model_3$`Pr(>F)` , scientific = F)
```

```
## [1] "          NA" "0.000000001550495"
```

The above p-value of the Model 1 and 2 ANOVA indicates it is much smaller than 0.05 and therefore we cannot quantify our results over one model from the other and both models are relatively the same.

Conclusion

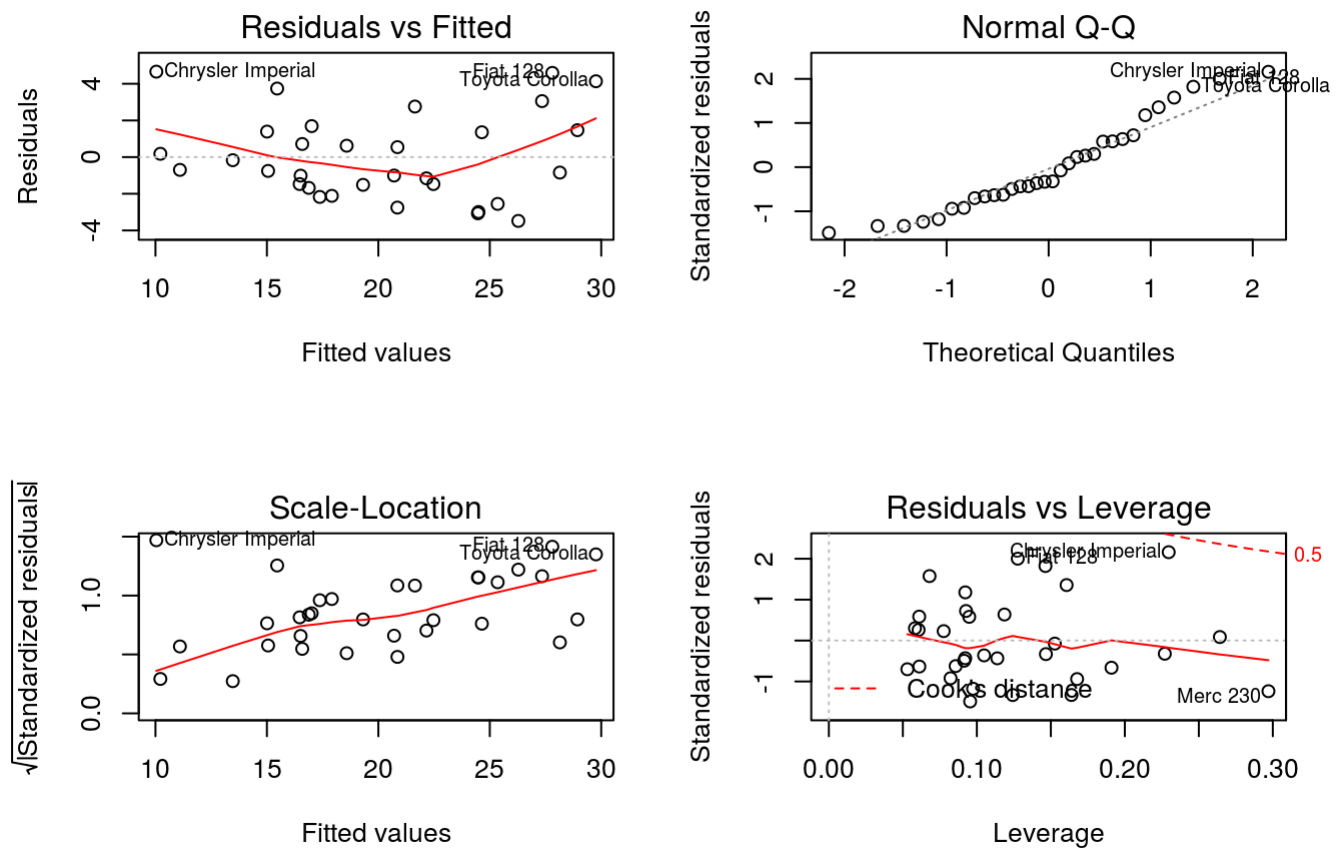
```
Model_2$coefficients
```

```
## (Intercept)          wt          qsec    amManual
##    9.617781    -3.916504    1.225886    2.935837
```

In conclusion, holding the weight and acceleration (qsec) of the car constant (shown in Model 2's coefficients above), manual transmission cars offer 2.94 MPG better fuel efficiency.

Appendix

```
par(mfrow=c(2,2))
plot(Model_2)
```



Per the above plot of the residuals, we observe there are a few outliers and there is nothing significant that skews the data.

Works Cited

¹ Henderson and Velleman (1981), Building multiple

² R-bloggers; Arnold Salvacion, 2014-11-13 (<https://www.r-bloggers.com/scatter-plot-matrices-in-r/>)