MT Cars Analysis

Jayesh Gokhale

5/1/2021

Analysis on Motor Trend Cars Dataset

Executive Summary

The data set is based on 1974 Motor Trend US magazine and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles.

We have to answer two questions -

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

Numeric variables - **mpg (Miles/(US) gallon)**, disp (Displacement), hp (Gross Horsepower), drat (Gear Axle Ratio), wt (Weight in 1000 lbs) and qsec (1/4 mile time)

Categorical Variables - cyl (Number of cylinders), vs (Engine: 0 = v-shaped, 1 = straight), **am (Transmission: 0 = automatic, 1 = manual)**, gear (Number of forward gears), carb (Number of Carburetors)

BoxPlot Observations (Box Plot in Appendix)

1. Manual transmission does appear to have significantly higher MPG across Engines, Carburetors and No of forward gears wherever applicable
2. For 4 cylinder engines, manual transmission seems to give higher MPG
3. **In 6 & 8 cylinder engines, it seems to be too close to call.**

Heatmap Observations (Heatmap in Appendix)

1. We can see that wt, disp and hp are negatively highly correlated to mpg
2. drat is positively correlated to mpg.
3. We can also observe that wt, disp, hp and drat are correlated amongst themselves (either + or -).

Let us now look at the relationship between am (Transmission) and mpg while adjusting for each of the four numeric variables viz. wt, disp, hp and drat. The adjustment Plot is in the appendix.

Adjustment Plot Observations (Adjustment Plot in Appendix)

1. Except horsepower, Weight, Displacement and Gear Axle Ratio explains for a lot of variation in MPG. So does Displacement and Gear Axle Ratio.
2. We need to do here is a residual analysis. We need to study the how much of the residual variation in MPG is explained by Transmission after removing the effect of weight, hp, disp and drat one by one.

Before residual analysis let us do a quick ANOVA to see whether we should add no of cylinders in our model or not since that does not strictly follow the fixed pattern.

```
fit1 <- lm(mpg~wt,data=mt)
fit2 <- update(fit1,mpg~wt+cyl)
fit3 <- update(fit1,mpg~wt*cyl)
anova(fit1,fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + cyl
## Model 3: mpg ~ wt + cyl + wt:cyl
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     30 278.32
## 2     28 183.06  2    95.263 7.9443 0.00203 **
## 3     26 155.89  2    27.170 2.2658 0.12386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova Observation

Thus we can see that cylinder is significant here and we will add it to our residual analysis for weight. Let us add it for others too (but not its interaction as is evident from figures above).

Residual Plot Observations (Residual Plot in Appendix)

All the residual plots are homoscedastic. That is I cannot make out any pattern from these plots. This means that if we take example of weight (first plot), ***If the effect of weight and number of cylinders is removed, merely the fact that an engine is automatic or manual does not seem to have any impact on the MPG. The same holds true for effect of Displacement, HP and Gear Axle Ratio combined with No. of Cylinders***

Question 2 is "Quantify the MPG difference between automatic and manual transmissions"

If we do not account for any other factors then it is the difference of means + or - the pooled variance. Let us calculate that.

```
getTStats <- function(x1,x2,printLabel)
{
  s1 <- sd(x1); s2 = sd(x2); n1 <- length(x1); n2 <- length(x2);
  s <- sqrt(((n1-1)*(s1**2) + (n2-1)*(s2**2)) / (n1+n2-2))
  mu1 <- mean(x1); mu2 <- mean(x2)
  diffMU <- mu2 - mu1
  t.statistic <- diffMU / (s * sqrt((1/n1)+(1/n2)))
```

```
  p.value <- pt(t.statistic,n1+n2-2,lower.tail = FALSE)

  print(paste0(printLabel,": ",round(diffMU,2)))

  return(p.value)

}
x1 <- mt[mt$am=="automatic",]$mpg; x2 <- mt[mt$am=="manual",]$mpg
x1.adj <- mt[mt$am=="automatic",]$resid.mpg.wt; x2.adj <- mt[mt$am=="manual",]$resid.mpg.wt

p.value <- getTStats(x1,x2,"Difference in Means")
```

```
## [1] "Difference in Means: 7.24"
```

```
p.value.adj <- getTStats(x1.adj,x2.adj,"Difference in Means on Adjustment")
```

```
## [1] "Difference in Means on Adjustment: 0.08"
```

If we do not account for any other variable, the difference between means of MPG amongst manual and automatic transmission is significant to a p-value of 1.4251037×10^{-4}.

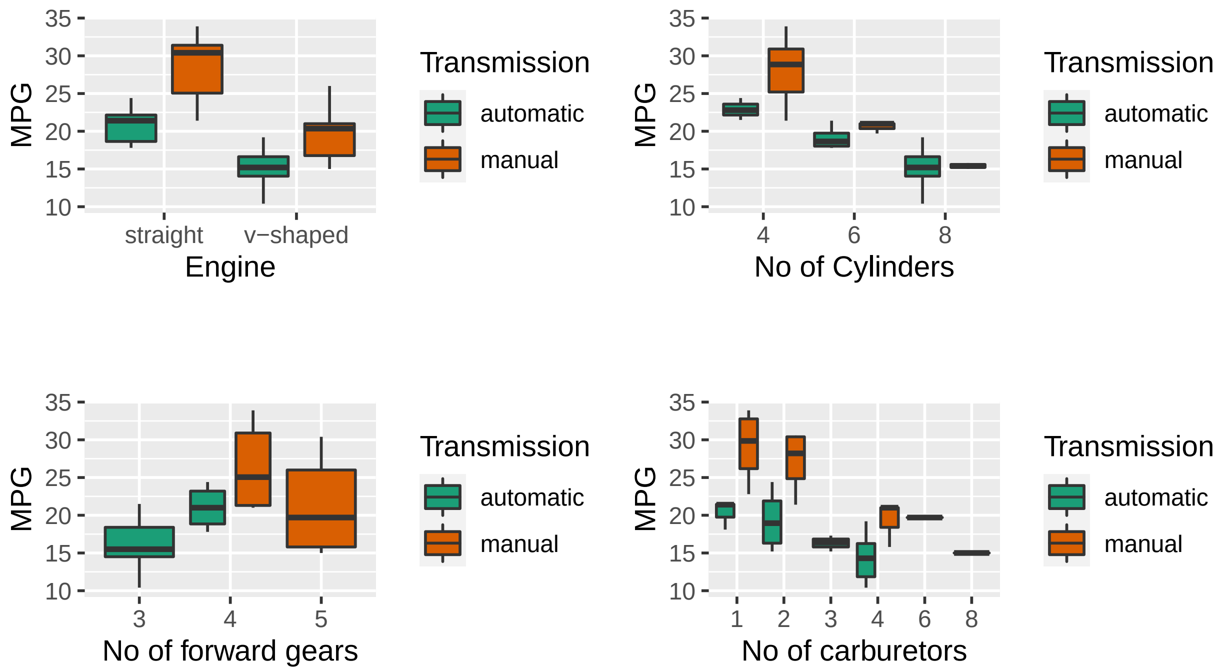
If we account for other variables like weight and no of cylinders, the difference between means of MPG amongst manual and automatic transmission is insignificant to a p-value of **0.4653558**.

Conclusion

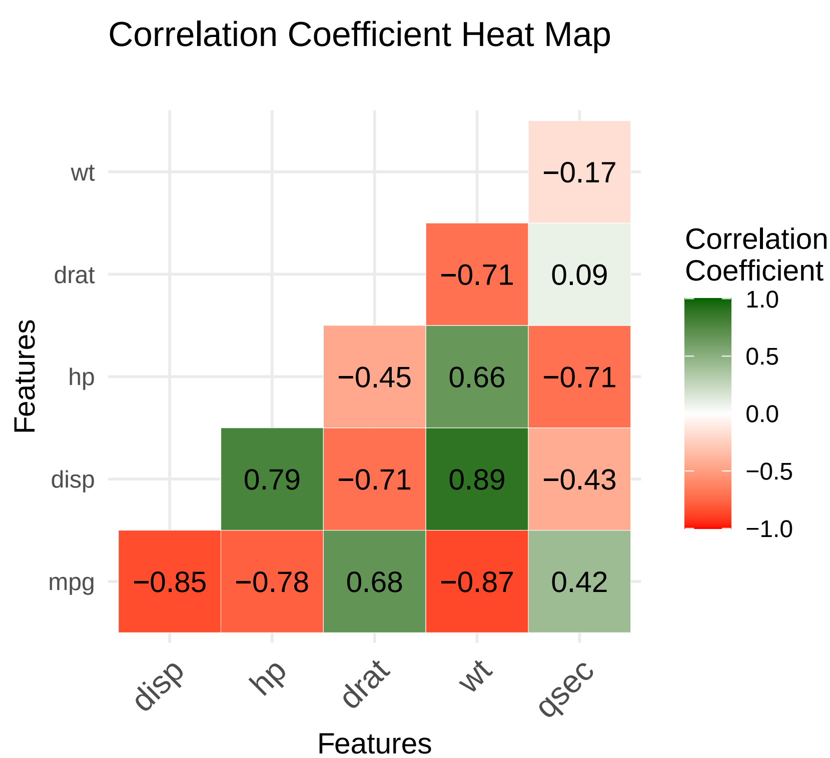
1. On the face of it it does appear that Transmission is highly correlated with MPG
2. However there are other factors like weight, displacement, gear axle ratio and number of cylinders which also seem to impact MPG. They are also correlated with each other
3. When the effect of such other factors are adjusted along with number of cylinders for eg- weight + cylinders, we cannot come to a conclusion that there is any residual correlation between MPG and Transmission (i.e. Automatic or Manual)
4. **Hence Question #1 cannot be answered with the given data**
5. **Consequentially Question #2 implies that there is no statistically significant difference between MPG for Automatic and Manual Transmission Engines by mere virtue of their transmission type (after adjusting for other factors)**

Appendix

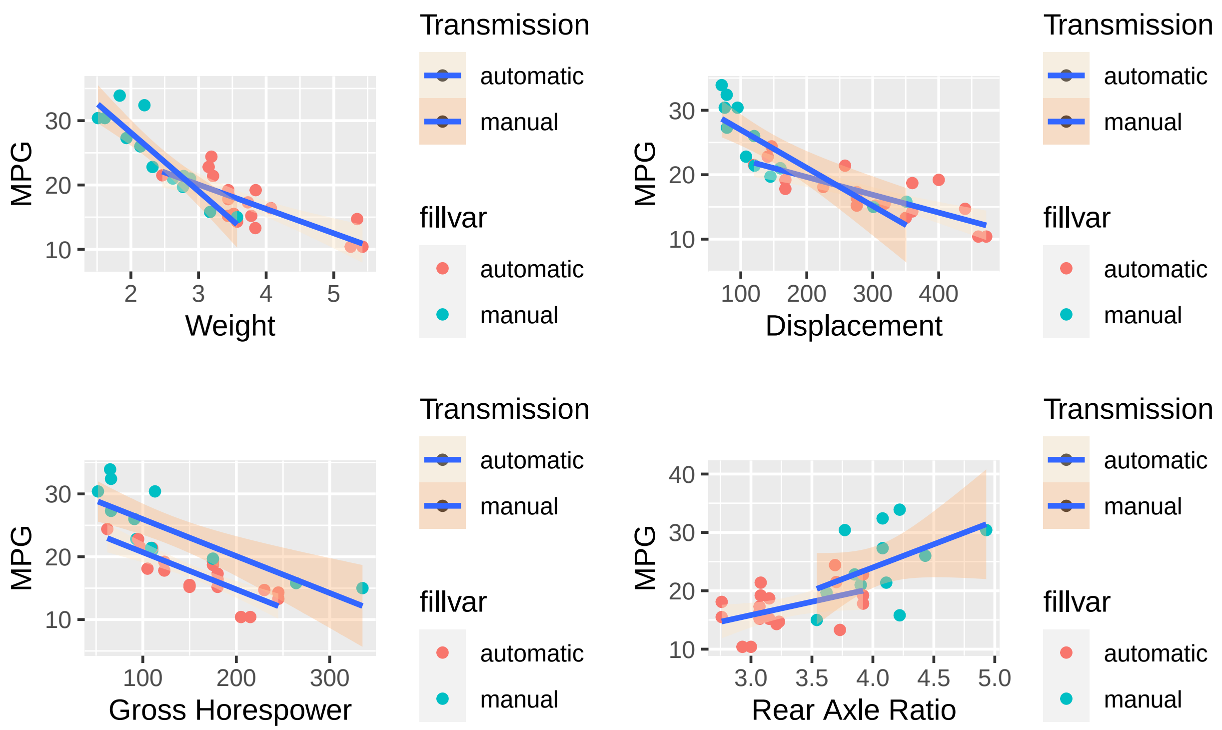
Box Plot



Heat Map



Adjustment Plots



Residual Plot