# Motor Trends Magazine - Impact of Transmission on MPG

Mark Wan

11/14/2020

## Executive Summary

---

This report presents an analysis of the impact of transmission type on fuel efficiency. Specifically, this report addresses the questions,

1. Does fuel efficiency, measured in terms of miles per gallon (MPG) differ between vehicles with auto transmission compared to vehicles with manual transmission?
2. If so, what is the difference in MPG?

### Load Data

---

```
data(mtcars)
mtcars2 <- mtcars
```

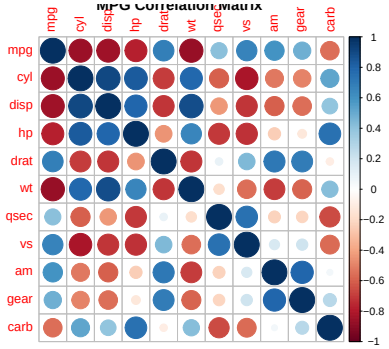
### Data Exploration

---

Appendix C presents a summary of the data.

The correlation matrix below indicate that numerous variables in the data set are correlated together. This observation is backed up by the paired plots (Appendix A) which lack randomness in many of the paired plots. This result suggests possible bias when examining regressors in isolation. Furthermore, the data shows that transmission (column "am") is highly correlated with drat, wt and gear. Hence, one should be careful to avoid regressing mpg vs am in isolation.

```
mtcars2.cor <- cor(mtcars2)
corrplot(mtcars2.cor, main="MPG Correlation Matrix")
```




In addition, the correlation maatrix of mpg vs all regressors suggests that transmission may not be the only driver for mpg.

```
cor(mtcars2$mpg,mtcars2[,-1])
```

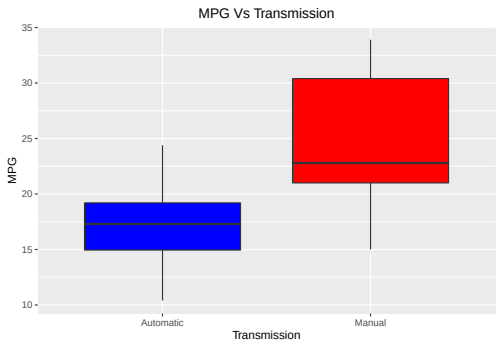
```
##            cyl       disp         hp      drat         wt     qsec        vs
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684 0.6640389
##             am      gear       carb
## [1,] 0.5998324 0.4802848 -0.5509251
```

The code below coverts transmission and engine shape in factor variables. This conversion is required for regression analysis.

```
fac_var <- c(8,9)
mtcars2[,fac_var] <- lapply(mtcars2[,fac_var],as.factor)
levels(mtcars2$am) <- c("Automatic", "Manual")
levels(mtcars2$vs) <- c("vshape", "straight")
```

The plot below indicates an increase in mpg moving from auto to manual cars. However, this observation needs to be validated with consideration to other regressors.

```
g <- ggplot(mtcars2,aes(x=am,y=mpg)) + geom_boxplot(fill=c('blue','red')) +
  labs(x="Transmission",y="MPG",
       title="MPG Vs Transmission") +
  theme(plot.title = element_text(hjust = 0.5))
g
```



**Results**

---

```
fit1 <- lm(mpg~am,data=mtcars2)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars2)
##
## 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 0.00000000000000113 ***
## amManual       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
```

The regression of mpg vs transmission (above) indicates an increase of 7.245 in mpg moving from auto to manual. The p-value of 0.000285 is also less than 0.05 significance level, indicating that the difference is signifcant. But how much of that difference can really be attributed to transmission? The following section attempts to answer that question.

```
ttest <- t.test(mtcars2$mpg~mtcars2$am,conf.level = 0.95)
ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars2$mpg by mtcars2$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 Automatic    mean in group Manual
##                17.14737                24.39231
```

The t-test above shows that the p-value (0.0013736) for difference in mean mpg for auto and manual vehicles is less than 0.05 significance level. Hence, the difference in mean mpg between auto and manual vehicles is significant and reinforces the point made in the previous regression analysis.

```r
fit_all <- lm(mpg~.,data=mtcars2)
summary(fit_all)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vsstraight   0.31776    2.10451   0.151   0.8814
## amManual     2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 0.0000003793
```

Interestingly, the regression of mpg against all variables (above) shows that the difference in mpg as a result of transmission (holding other regressors constant) is now insignificant. Though note that using all variables may not be appropriate if the variables are not all orthogonal (independent). In this case, the model may be overfitted with an excessive number of regressors, reducing the number of degrees of freedom that the model is allowed to arrive at the optimum solution.

```r
fit_optimum <- step(lm(mpg~.,data=mtcars2),trace=0)
summary(fit_optimum)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars2)
##
## 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 0.00000695 ***
```

```
## 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: 0.0000000000121
```

Using the step function to select the optimum model (bias vs variance), the optimum model shows that mpg is not only related to transmission but also to weight and qsec. Specifically, the model indicates that 1) a 1000lbs increase in wt results in a -3.9165037 increase in mpg, 2) a 1 sec increase in 1/4 mile time results in a 1.225886 increase in mpg, and 3) shifting from auto to manual results in a 2.9358372, with an error of 1.4109045, increase in mpg (holding other regressors constant). In addition, these differences are all significant as indicated by the p-values ($< 0.05$ significance level).

```
shapiro.test(fit_optimum$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit_optimum$residuals
## W = 0.9411, p-value = 0.08043
```

Finally, the P-value from a Shapiro-Wilk normality test is greater than 0.05, indicating that the assumption of normal residuals is reasonable (for a 0.05 significance level).

Residual and normal QQ plots in Appendix B also supports this assertion.

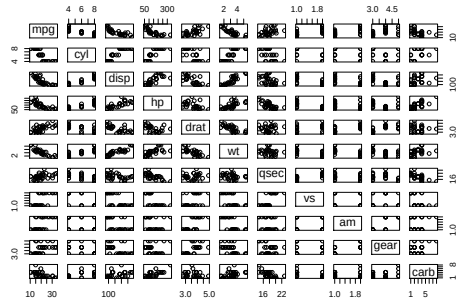
## Conclusion

In conclusion, the results of the multi-variable regression indicate that transmission does have an impact on mpg; manual cars do have, on average, higher mpg. However, in order to estimate the absolute mpg values, weight and 1/4 mile times must also be considered.

## Appendix A - Paired Plots

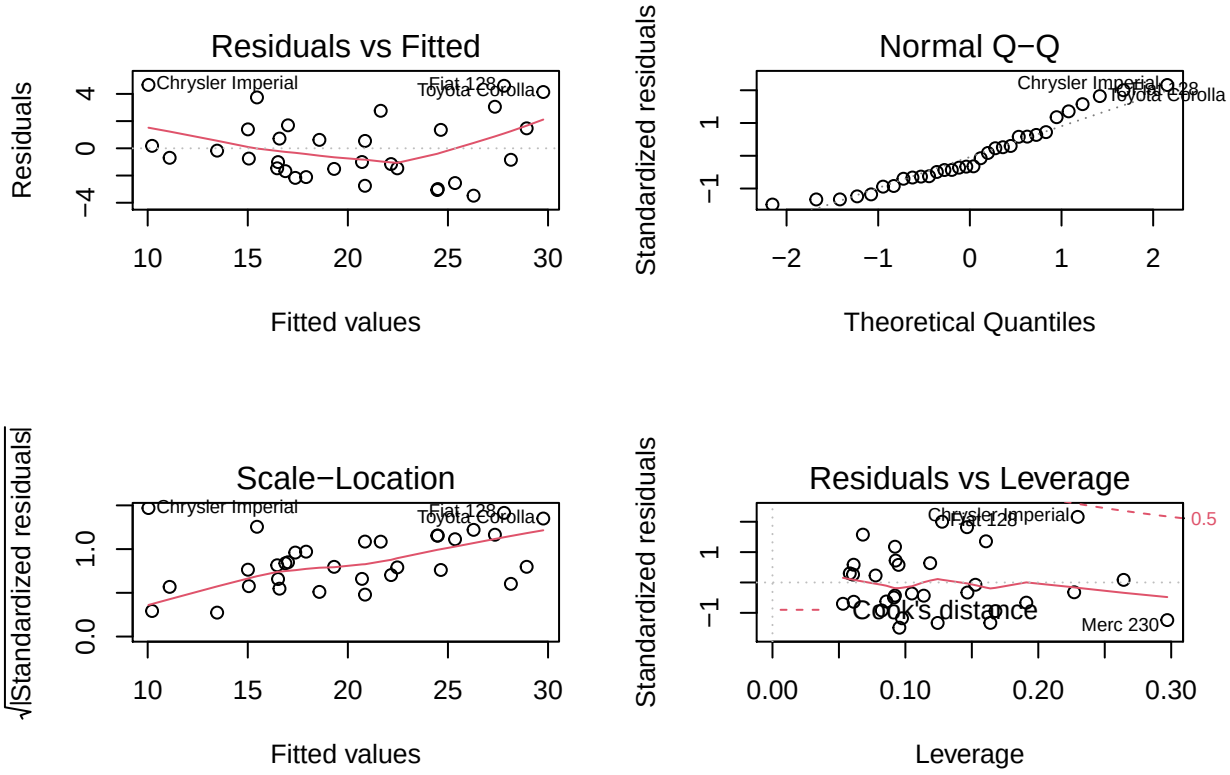
---

```
pairs(mtcars2)
```



## Appendix B - Residual Plots

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




## Appendix C - Summary of mtcars

```
summary(mtcars2)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs             am       
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   vshape  :18   Automatic:19  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   straight:14   Manual   :13  
##  Median :3.695   Median :3.325   Median :17.71                               
```

```
##  Mean   :3.597   Mean   :3.217   Mean   :17.85
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90
##  Max.   :4.930   Max.   :5.424   Max.   :22.90
##       gear            carb
##  Min.   :3.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:2.000
##  Median :4.000   Median :2.000
##  Mean   :3.688   Mean   :2.812
##  3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :8.000
```