# MPG Analysis on Auto vs Manual Transmissions

*M Erlandson*

*July 9, 2016*

## Summary

This is a data analysis of the mtcars dataset included in R. We will be analyzing how much several variables, such as the number of cylinders or carborators, affect the MPG outcome on both standard (manual) and automatic transmissions.

## Initial Exploration

We first load the data, all necessary libraries for analysis (ggplot2 and GGally), and take a quick look at the headers and first values. (Output from str(mtcars) hidden to shorten this paper.)

From the write up, We know that cyl, vs, am, gear, and carb can all be factor variables, and it is the AM variable (automatic 0 or manual 1 transmission) that is the focus of this study. Appendix plot #1 is a graph of all the variables cross-plotted with each other to see what relationships we might have. There is quite a bit of correlation between most of the variables and mpg (shown all in column 1), with the exception of qsec, which doesn't make sense in relation to mpg anyway.

## MPG Analysis

### Question 1) Which transmission is better for MPG?

```
fitAM <- lm(mpg ~ am, mtcars)
coef(fitAM)
```

Since mpg is a continuous value, we can use linear regression for our models. This simple linear regression says that there is a 7.24 mpg increase going from automatic to manual transmission, as seen in Appendix plot #2, but is there more going on here? What other factors might be involved?

### Question 2) Quantify the MPG difference between transmission types.

Let's look at comparisons with all the factor variables. We'll start with number of gears, since that's most related to the transmission, then add cylinders to the mix, and finally carborators that are attached to the cylinders in the engine. (See the appendix for R code and output.)

The comparison shows that the inclusion of gears does not make any significant change in the analysis, but the inclusion of cylinders does in a very big way (and a negative impact, at that) and carborators to a smaller extent.

We can run a second test with the other non-factor variables, also shown in the appendix. According to those results, displacement has a huge effect on mpg (also negatively, so the bigger the engine, the less efficient it is), with hp and wt coming in a distant second and third, similar to carborators.

## Final Model

Let's fit a model with the two biggest adjustments, cyl and disp.

```
fitAMcylDisp <- lm(mpg ~ am + cyl + disp, mtcars)
summary(fitAMcylDisp)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 32.91685995  2.7791366 11.844276 2.026114e-12
## am           1.92872540  1.3397251  1.439643 1.610559e-01
## cyl         -1.61821911  0.6993650 -2.313841 2.823412e-02
## disp        -0.01559041  0.0106533 -1.463435 1.544849e-01
```

Now taking into account the other variables, the adjusted coefficient for AM is 1.93. This is still positive, so the manual transmission might give a better MPG performance (34.85 mpg) than the automatic (32.92 mpg), though it's p-value says there's not a significant difference.

## Residuals

Let's check our residuals to see if this is right. Appendix plot #3 shows the residuals after cyl has been removed and plot #4 shows the residuals after disp has been removed. In both cases, the values are much more random and don't appear to have any significant linear relationship. This agrees with our estimate that there is really no change between manual and automatic transmissions.

## Uncertainty

How much do the additional regressors change the variance? According to theory, *any* addition of regressors increases our standard error. The simple model has a small std error (1.1246025) whereas the more multivariate model has a larger std error (2.7791366), almost twice, and it makes us less certain of our outcomes.

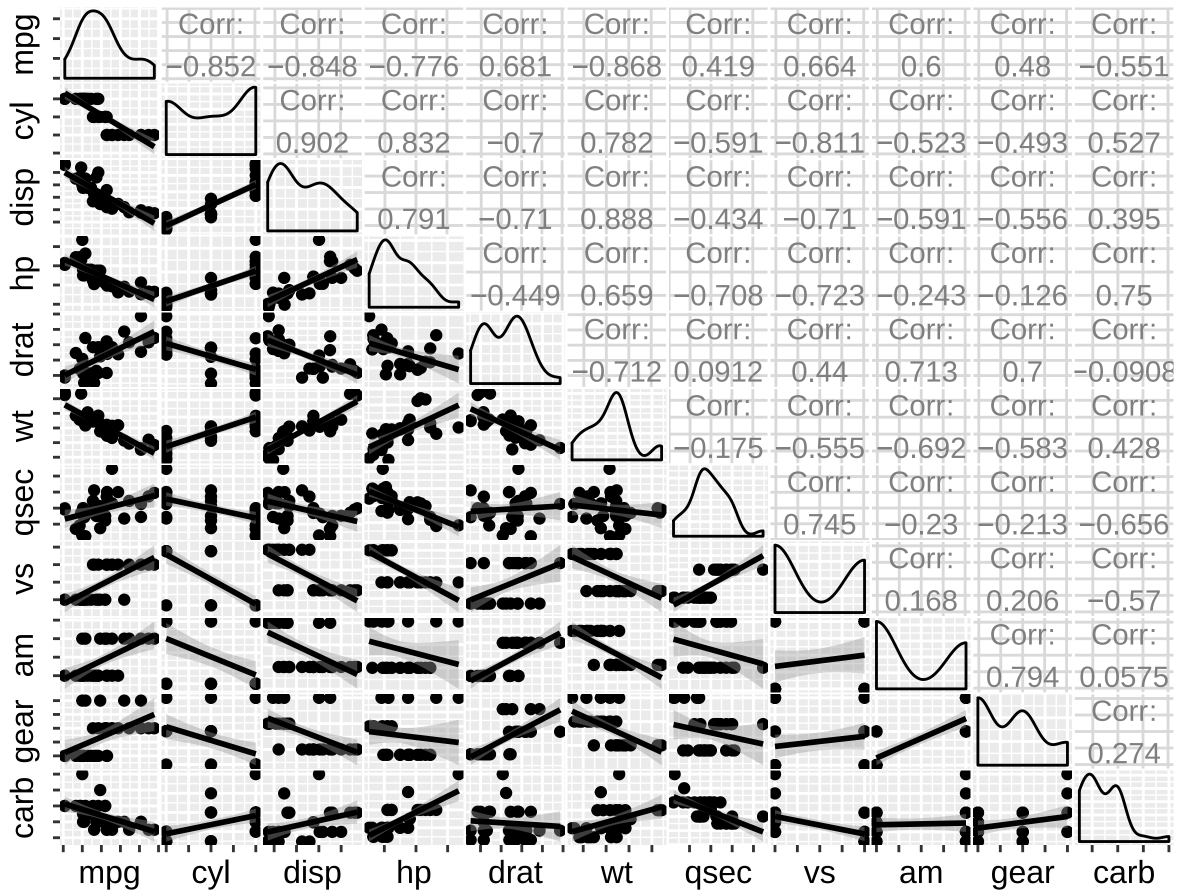
We should also check our assumptions using the gvlma pkg, shown below. According to the test, our global stat and link functions are not satisfied (p-values less than 5%) so a new analysis may be needed.

```
gvmodel <- gvlma(fitAMcylDisp)
display.gvlmatests(gvmodel)
```

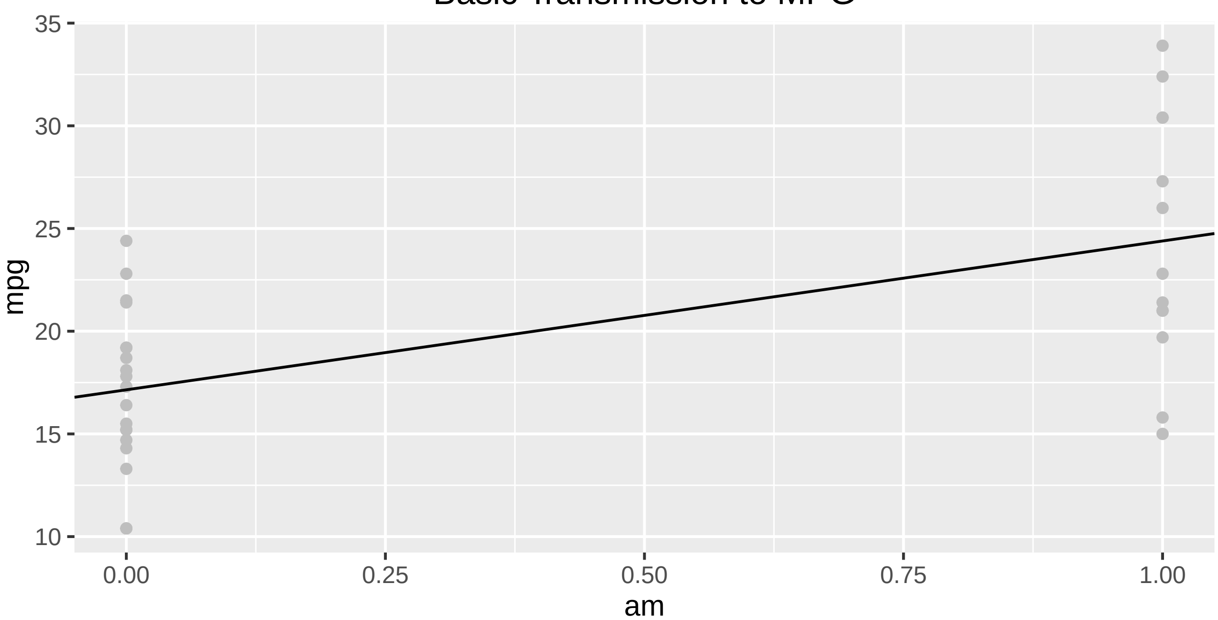
```
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fitAMcylDisp)
##
##                        Value  p-value                   Decision
## Global Stat        10.16421 0.037751 Assumptions NOT satisfied!
## Skewness            2.14475 0.143058    Assumptions acceptable.
## Kurtosis            0.06886 0.793007    Assumptions acceptable.
## Link Function       6.87272 0.008752 Assumptions NOT satisfied!
## Heteroscedasticity  1.07788 0.299173    Assumptions acceptable.
```

## Appendix

Cross−plot of Mtcars Variables




Basic Transmission to MPG



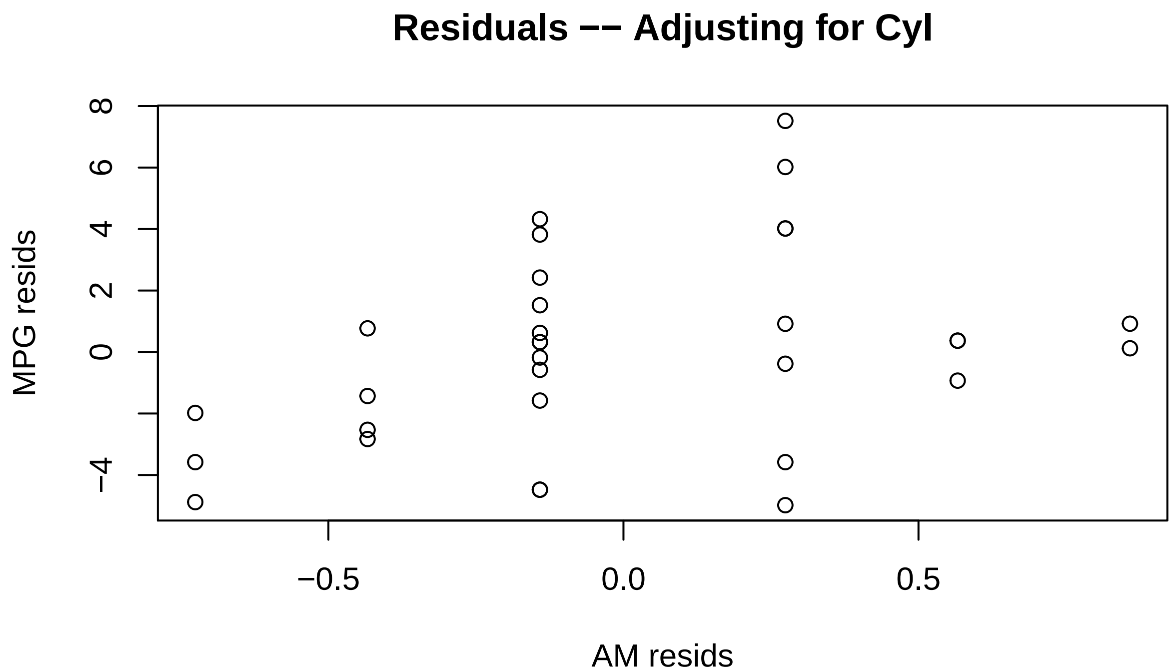
```r
fitAM1 <- update(fitAM, mpg ~ am + gear)
fitAM2 <- update(fitAM, mpg ~ am + gear + cyl)
fitAM3 <- update(fitAM, mpg ~ am + gear + cyl + carb)
anova(fitAM, fitAM1, fitAM2, fitAM3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + gear
## Model 3: mpg ~ am + gear + cyl
## Model 4: mpg ~ am + gear + cyl + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 720.85  1      0.05  0.0063   0.93723
## 3     28 262.40  1    458.45 59.9188 2.523e-08 ***
## 4     27 206.58  1     55.82  7.2961   0.01179 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fitAM4 <- update(fitAM, mpg ~ am + disp)
fitAM5 <- update(fitAM, mpg ~ am + disp + hp)
fitAM6 <- update(fitAM, mpg ~ am + disp + hp + drat)
fitAM7 <- update(fitAM, mpg ~ am + disp + hp + drat + wt)
anova(fitAM, fitAM4, fitAM5, fitAM6, fitAM7)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + disp
## Model 3: mpg ~ am + disp + hp
## Model 4: mpg ~ am + disp + hp + drat
## Model 5: mpg ~ am + disp + hp + drat + wt
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.90
## 2     29 300.28  1    420.62 62.2539 2.29e-08 ***
## 3     28 226.10  1     74.18 10.9788 0.002716 **
## 4     27 221.04  1      5.06  0.7491 0.394664
## 5     26 175.67  1     45.37  6.7157 0.015470 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(resid(lm(am~cyl, mtcars)), resid(lm(mpg~cyl, mtcars)),
     main="Residuals -- Adjusting for Cyl", xlab="AM resids", ylab="MPG resids")
```




```r
plot(resid(lm(am~disp, mtcars)), resid(lm(mpg~disp, mtcars)),
     main="Residuals -- Adjusting for Disp", xlab="AM resids", ylab="MPG resids")
```