Motor Trend Cars - How is MPG ratingaffected by transmission type and other specs?

David C Latshaw II

Executive Summary

We have built a linear model that takes into consideration the weight of the car, how many cylinders it has, it's horsepower, and its transmission type to predict the miles per gallon a car will get. The results show that if you want to maximize your the miles per gallon you should have a light weight car with 4 cylinders, lower horsepower, and a manual transmission.

Introduction

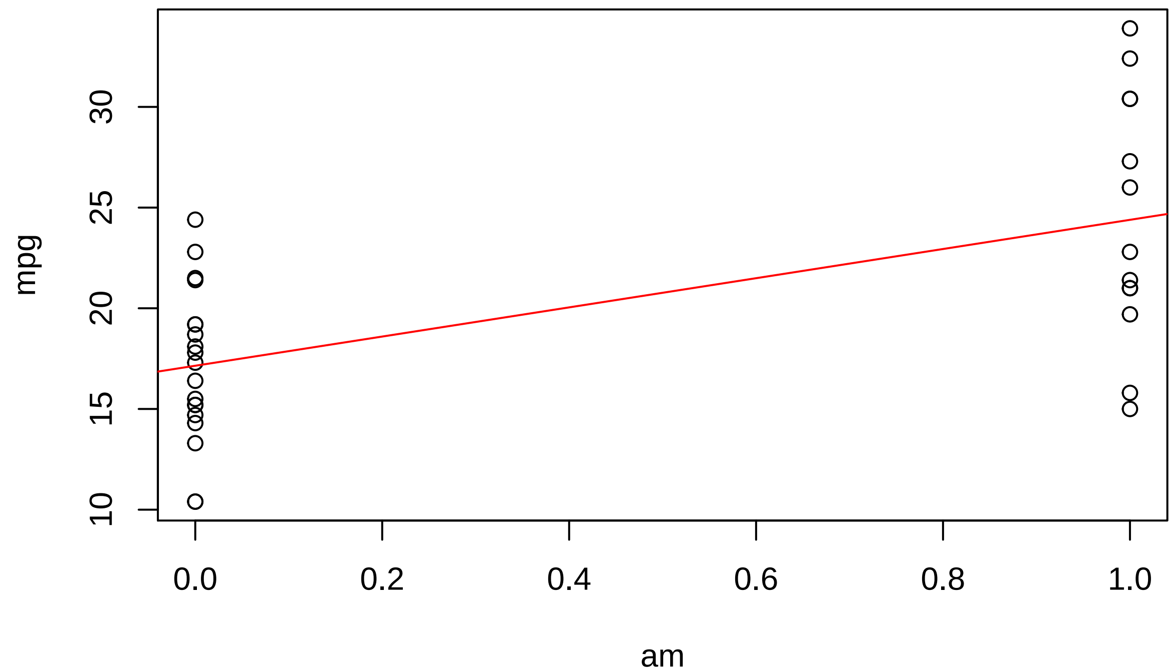
For this analysis we will use the Motor Trend Cars data set in R. Ultimately we would like to answer the questions:
Is an automatic or manual transmission better for MPG?
Quantify the MPG difference between automatic and manual transmissions?

Data First we load the load the data

```
data(mtcars)
```

Analysis Next we will take a look at a plot of the MPG distribution for automatic and manual transmissions with 0 being automatic and 1 being manual:

```
plot(mpg~am,data=mtcars)
abline(lm(mtcars$mpg~mtcars$am),col="red")
```



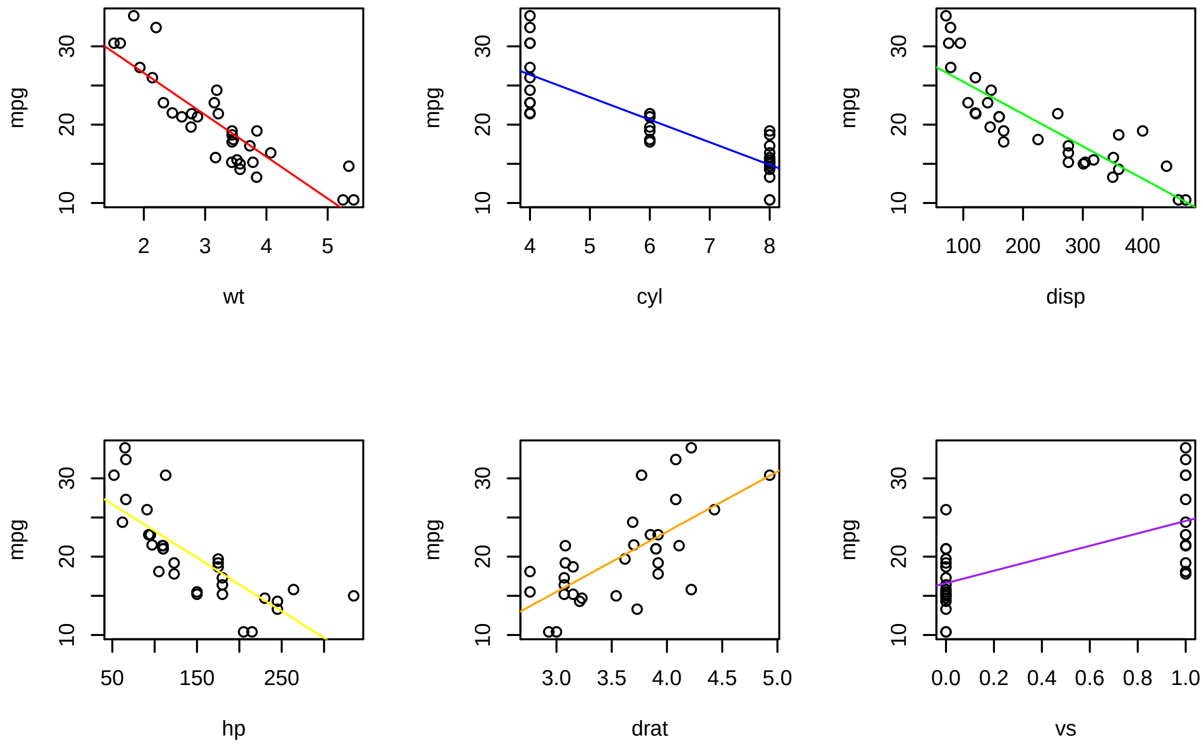
There appears to be a bit of a correlation between what type of transmission the car has and it's MPG rating, but we would also like to know how the other variables in the data set correlate with MPG:

```
cor(mtcars$mpg,mtcars)
```

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

From these correlations we will examine all of the variables that have a correlation higher than the transmission type to start:

```
par(mfrow=c(2,3))
plot(mpg~wt,data=mtcars)
abline(lm(mtcars$mpg~mtcars$wt),col="red")
plot(mpg~cyl,data=mtcars)
abline(lm(mtcars$mpg~mtcars$cyl),col="blue")
plot(mpg~disp,data=mtcars)
abline(lm(mtcars$mpg~mtcars$disp),col="green")
plot(mpg~hp,data=mtcars)
abline(lm(mtcars$mpg~mtcars$hp),col="yellow")
plot(mpg~drat,data=mtcars)
abline(lm(mtcars$mpg~mtcars$drat),col="orange")
plot(mpg~vs,data=mtcars)
abline(lm(mtcars$mpg~mtcars$vs),col="purple")
```



As a first pass we will include all of these variables as predictors in the model with MPG as the response:

```r
mtcars$am<-factor(mtcars$am)
mtcars$cyl<-factor(mtcars$cyl)
fit<-lm(mpg~wt+cyl+disp+hp+drat+vs+am,mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + disp + hp + drat + vs + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2107 -1.1566 -0.1668  1.2068  4.6460
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.82997    6.74447   4.423 0.000196 ***
## wt          -2.59462    1.20130  -2.160 0.041449 *
## cyl6        -2.05552    1.80311  -1.140 0.266024
## cyl8        -0.02330    3.81651  -0.006 0.995181
## disp         0.00436    0.01304   0.334 0.741057
## hp          -0.03580    0.01463  -2.446 0.022514 *
## drat         0.38814    1.46606   0.265 0.793559
## vs           2.00490    1.82995   1.096 0.284593
## am1          2.55899    1.74302   1.468 0.155612
```

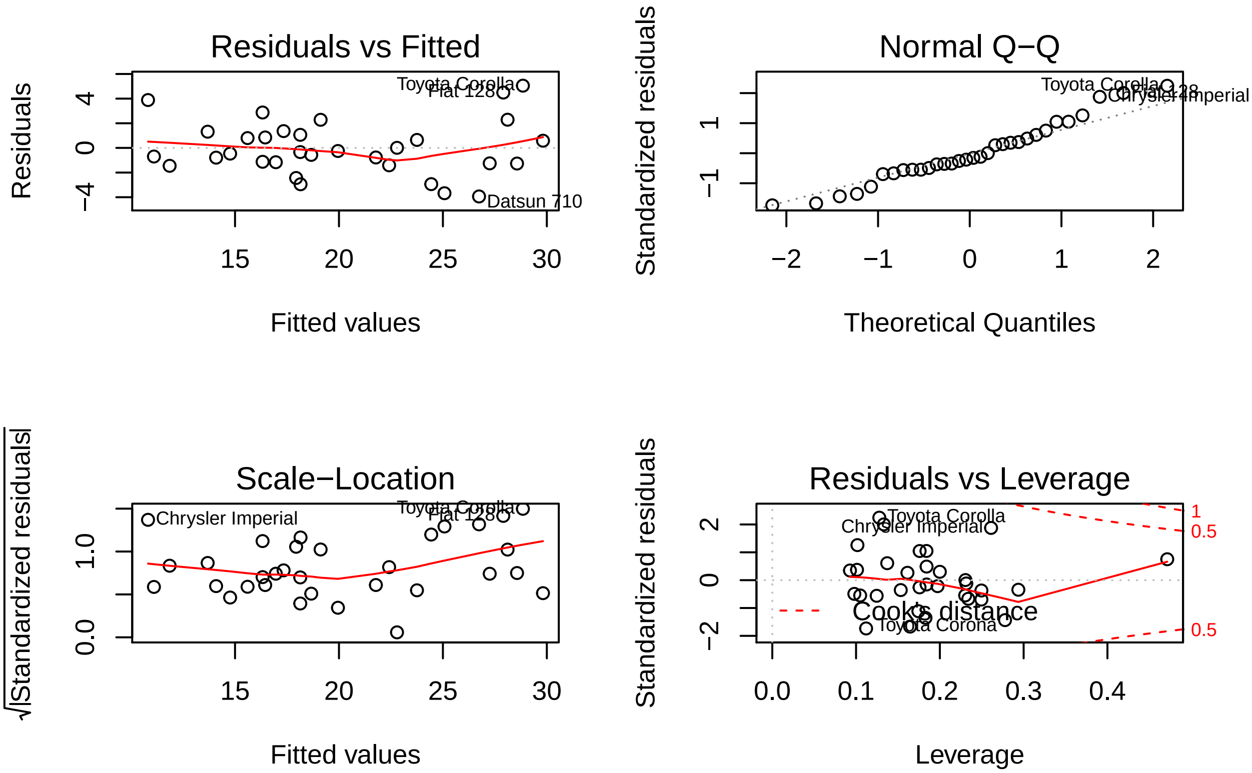
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.49 on 23 degrees of freedom
## Multiple R-squared:  0.8733, Adjusted R-squared:  0.8292
## F-statistic: 19.82 on 8 and 23 DF,  p-value: 1.351e-08
```

From the fit we can see that there are a few variables that we can get rid of to simplify the model because they have a P-value >0.05. We will start with the three lowest contributors, namely we will get rid of displacement (disp), the rear axle ratio (drat), and the engine configuration (vs) then rebuild the model.

```r
mtcars$am<-factor(mtcars$am)
mtcars$cyl<-factor(mtcars$cyl)
fit<-lm(mpg~wt+cyl+hp+am,mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + hp + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## wt          -2.49683    0.88559  -2.819  0.00908 **
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## am1          1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

```r
par(mfrow=c(2,2))
plot(fit)
```



In this final model we have included transmission type (am) as well as the other highest contributing variables: weight (wt), number of cylinders (cyl), and horsepower (hp). Overall this model can describe 86.6% of the variability in MPG between cars. The variables that are most highly correlated to MPG in the model are the weight and horsepower of the car. In this particular model transmission type does not influence the MPG rating (p-value > 0.05). A review of the residuals also shows that the distribution of the data in the model is normal and not dependent on what variables were chosen for the fit.

Results Some of the most important results from the model are as follows:

- For every 1000lb increase in the weight of the car there is a 2.50 mile per gallon decrease
- When compared to 4 cylinder cars, 6 cylinder cars get 3.03 less miles per gallon and 8 cylinder cars get 2.16 less miles per gallon (uncertain with a p-value > 0.05)
- A 1hp increase correlates to a decrease of 0.032 miles per gallon
- Using a manual transmission instead of an automatic transmission (uncertain with a p-value > 0.05)

Taking these general trends into consideration, if you want to maximize your the miles per gallon you should have a light weight car with 4 cylinders, lower horsepower, and a manual transmission.