Manual vs Automatic Transmission

Kyle Scully

November 22, 2014

Executive Summary

A common question motor enthusiast often ask is does transmission type have significant impact on fuel efficiency. We will analyze a dataset from Motor Trend magazine which comprises fuel consumption and 10 aspects of automobiles to address these questions:

- Is an automatic or manual transmission better for MPG?
- If so, what is the quantity difference?

This report will show that using hypothesis testing and linear regression, manual transmissions are generally more fuel efficient than automatic transmissions. Taking confounding variables into account, we will use analysis of variance and multivariate regression to show that manual transmissions, on average, get 2 miles per gallon better fuel efficiency than automatic transmissions.

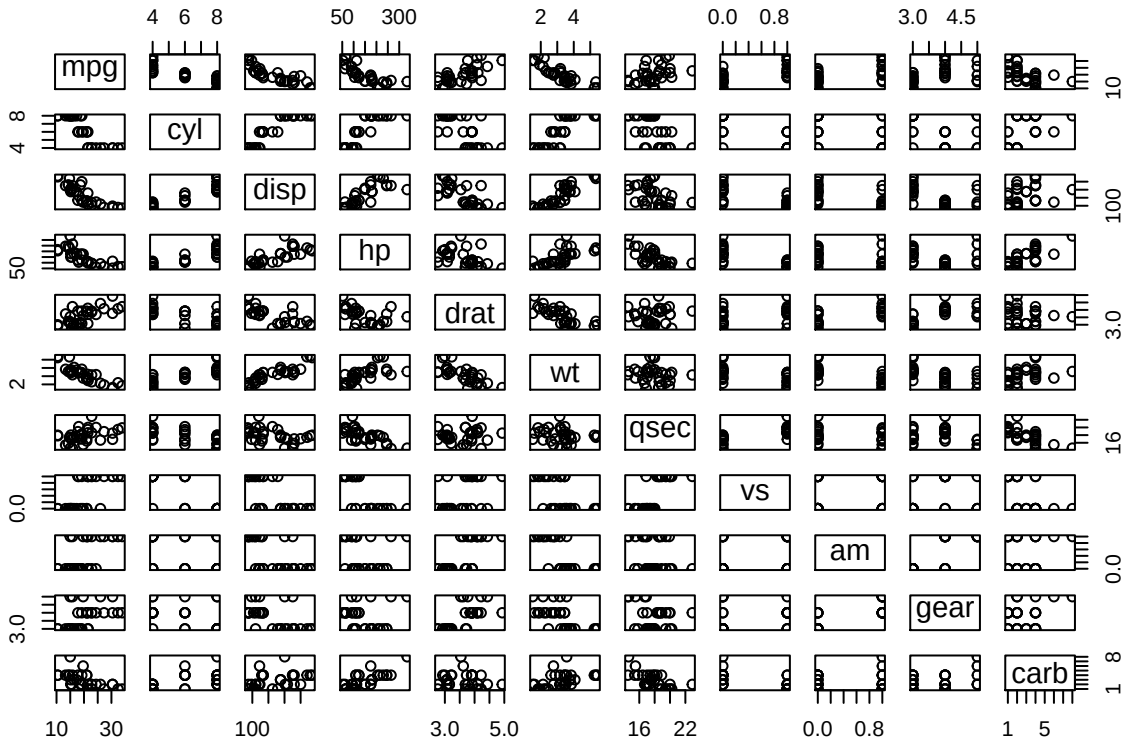
Exploratory Analysis

The Data Set

A data frame with 32 observations on 11 variables.

- [, 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 (lb/1000)
- [, 7] qsec 1/4 mile time
- [, 8] vs V/S
- [, 9] am Transmission (0 = automatic, 1 = manual)
- [,10] gear Number of forward gears
- [,11] carb Number of carburetors

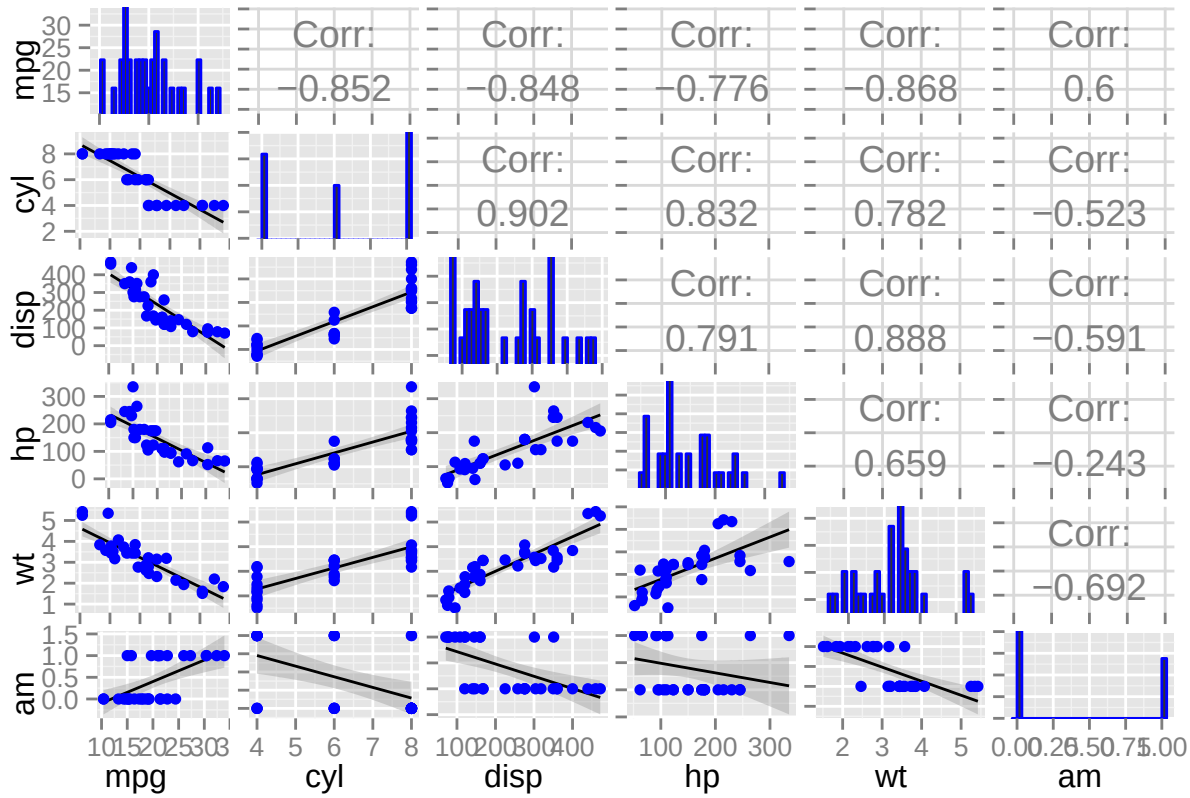
General Relation of Variables



From the above plot we can generally see how data variables affect MPG.

Focused Subset of Variables

Let's take a quick look at the distributions & the correlations between some of the variables that stand out:



From the above plot we can note a few things:

- The distribution of MPG is fairly normal
- There seems to be a higher average MPG for vehicles with manual transmissions
- There is a high correlation between mpg and wt, cyl, disp, and hp
- We also see a very high correlation between cyl and disp

```
means <- aggregate(mpg~am, data = mtcars, mean)
means
```

```
##   am      mpg
## 1  0 17.14737
## 2  1 24.39231
```

```
means[2,2] -means[1,2]
```

```
## [1] 7.244939
```

We can see that the average difference in mpg between transmission type is 7.2449393

Hypothesis Testing

We need to test the validity of this by running a t-test o nthe null-hypothesis:

```r
manual <- mtcars[mtcars$am == 1,]
auto <- mtcars[mtcars$am == 0,]
t.test(auto$mpg, manual$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  auto$mpg and manual$mpg
## 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 of x mean of y
##  17.14737  24.39231
```

From this p-value we must reject the null hypothesis.

This means that there is a significant difference in fuel efficiency with respects to transmission type: vehicles with manual transmissions on average have higher fuel efficiency than vehicles with automatic transmissions.

Linear Regression Model

Simple Linear Regression

Let's look at a simple linear regression between transmission type and miles per gallons:

```r
fit <- lm(mpg~am, data = mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## 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 1.13e-15 ***
## am             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
```

Looking at the Multiple R-squared we see this model only explains about 36% of the variance so it is not a good model

Multivariate Linear Regression

Earlier we noticed that cylinders and displacement were highly correlated to each other. We also noticed that horsepower has a high correlation to displacement. Due to these relationships we will exclude cylinders and displacement from being predictors in our model as they are collinear and more or less discrete values.

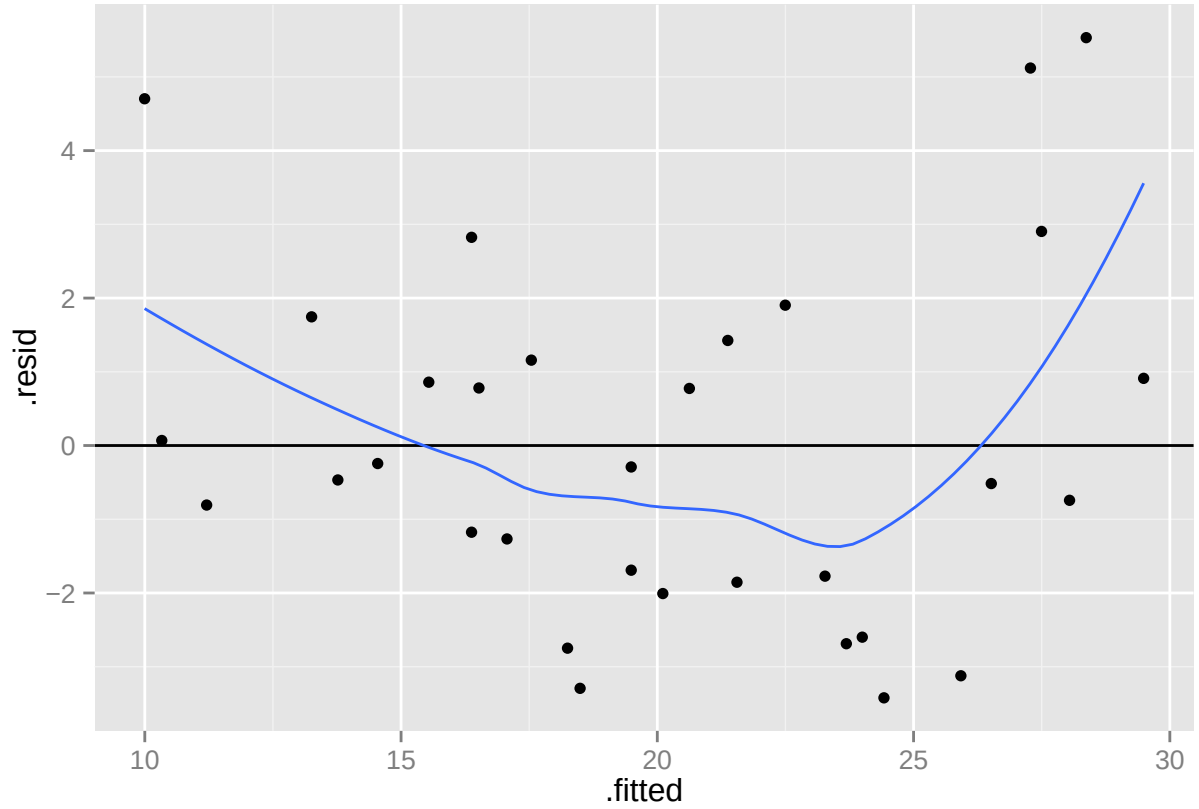
```
fit2 <- lm(mpg~am + wt + hp, data = mtcars)
anova(fit, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + hp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 180.29  2    540.61 41.979 3.745e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this p-value we must reject the null hypothesis.

This means that our multivariate model is significantly different from our simple model.

Residuals Analysis



From the plot above, the residuals are normally distributed.

Conclusions

```
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

We have concluded that **vehicles with manual transmissions, on average, have better (higher) MPG**.

From the figure above we can see that **vehicles with manual transmissions average roughly 2 miles per gallon more than vehicles with automatic transmissions**.

Additionally, we can see that weight had the largest confounding effect on the relationship between transmission type and MPG. Also, we can see from the Multiple R-squared that means in terms of uncertainty that this model explains about 84% of the variance.