Transmission Type and Fuel Efficiency

Taiki Sakai

October 24, 2015

Summary

In this paper we examine the effect of transmition type (automatic or manual) on fuel efficiency (miles per gallon). We will try to create a model for fuel efficiency using any number of variables among weight, horsepower, transmission type, and other data recorded for 32 different cars. After considering a few differnet models, we arrive at a model using horsepower and transmission type and conclude that an automatic transmission has lower fuel efficiency than a manual transmission

Exploratory Analysis

We start our analysis by loading in the dataset and taking a quick look at the correlation matrix and the pairs plot (Figure 3 in the Appendix). Only a subset of the correlation matrix is reproduced, for brevity.

```
##             mpg        cyl       disp         hp         wt         am
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684 -0.8676594  0.5998324
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475  0.7824958 -0.5226070
## disp -0.8475514  0.9020329  1.0000000  0.7909486  0.8879799 -0.5912270
## hp   -0.7761684  0.8324475  0.7909486  1.0000000  0.6587479 -0.2432043
## wt   -0.8676594  0.7824958  0.8879799  0.6587479  1.0000000 -0.6924953
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043 -0.6924953  1.0000000
```

From the correlation matrix we can see that horsepower, cylinder count, displacement, and weight are all highly correlated with each other. This makes sense, as a higher displacement engine would require more cylinders and should yield a higher horsepower. A higher weight vehicle will require more horsepower to move it effectively. Intuitively, it would make sense if one of (or all of) these variables would have an effect on fuel efficiency, and we can see thta they all have high correlation with mpg. Since they are all strongly correlated with each other, it may be possible to use just one of them as a proxy for all of the others as they generally measure a similar quality of a vehicle. The rest of the variables have a lower correlation with mpg, so these seem like a good place to start our analysis.

We also notice that cyl, disp, and wt all have a relatively high correlation with am, the other variable we are interested in. Because of this, we will begin our analysis using horsepower and transmission type, as they are both correlated with mpg but not strongly correlated with each other.

Model Building

We will start our model at a very basic level. There are clearly factors other than transmission type that should affect fuel efficiency, so we will not consider a model that only includes transmission type. Our first model will use horsepower and transmission type, and we will try to build more complex models from there.

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
## hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
## am           5.2770853 1.079540576  4.888270 3.460318e-05
```

With this model we can see that the coefficients for the intercept, hp, and am are statistically significant. The R-squared is fairly high at .767, suggesting the model is a decent fit for our data. The residual plot (Figure 1 in Appendix) does not seem to have any pattern to it, suggesting that the model might be reasonable.

From here we will try adding other variables from our dataset and use an anova test to determine if the additional variables contribute to the model. We will include all of the remaining variables in the dataset, and also an interaction variable between horsepower and transmission type. For brevity, the results of these are not shown.

These tests reveal that only one of the additional variables appears to significantly improve our model. That p-value corresponds to the weight variable. With this new knowledge we construct a model with horsepower, weight, and transmission type.

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
## am           2.08371013 1.376420152  1.513862 1.412682e-01
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
```

One thing to notice is that the additon of the weight variable has increased the standard error of the coefficient for our transmission variable to the point where it is no longer statistically significant. This could suggest that there is not a relationship between fuel efficiency and transmission type. However, we also notice from the residual plot (Figure 2 in Appendix) that there might be a slight negative or U-shaped pattern in the residuals. This suggests that the model might not be correct. With this concern in mind, we will choose to revert to our simpler model using only horsepower and transmission type.

Results

Our final model is:

```
##
## Call:
## lm(formula = mpg ~ hp + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3843 -2.2642  0.1366  1.6968  5.8657
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.584914   1.425094  18.655  < 2e-16 ***
## hp          -0.058888   0.007857  -7.495 2.92e-08 ***
## am           5.277085   1.079541   4.888 3.46e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.909 on 29 degrees of freedom
## Multiple R-squared:  0.782,  Adjusted R-squared:  0.767
## F-statistic: 52.02 on 2 and 29 DF,  p-value: 2.55e-10
```

These coefficients can be interpreted as: for every 1 increase in horsepower, mpg decreases by .0589. If a transmission is manual (am=1), mpg increases by 5.277. Thus we conclude that manual transmissions have higher fuel efficiency, since the coefficient 5.277 is positive and is statistically significant. A 95% confidence interval for the increased mpg in a manual transmission is 3.0476997, 7.5063003.

Appendix

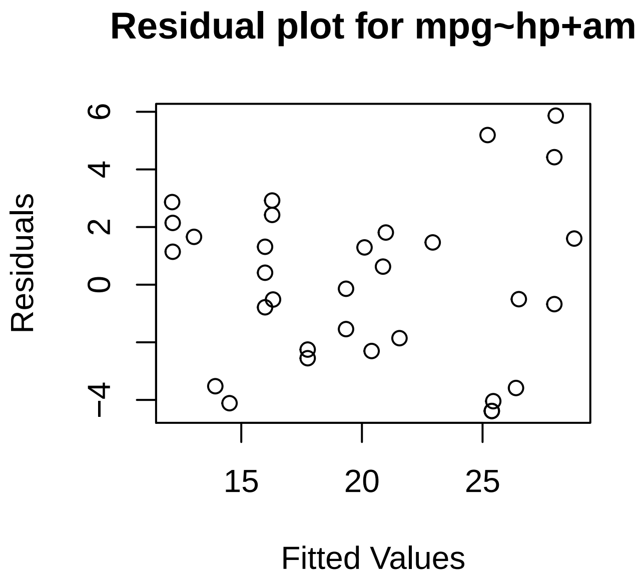

Figure 1: Residuals for Simpler Model

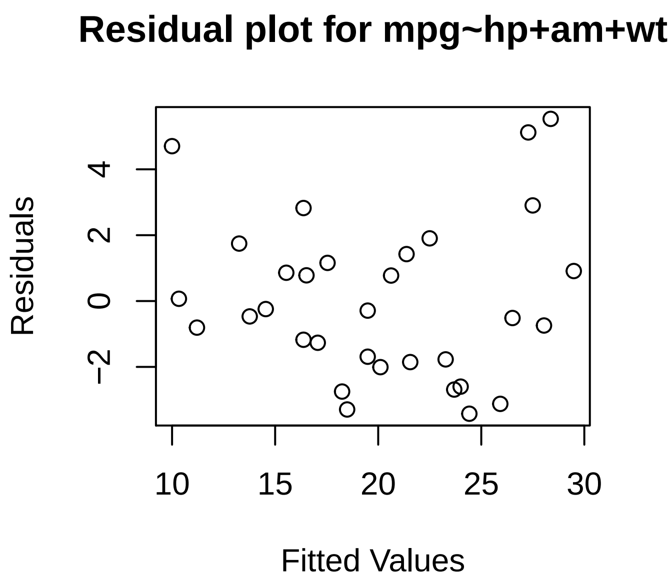

Figure 2: Residuals for Model with Weight

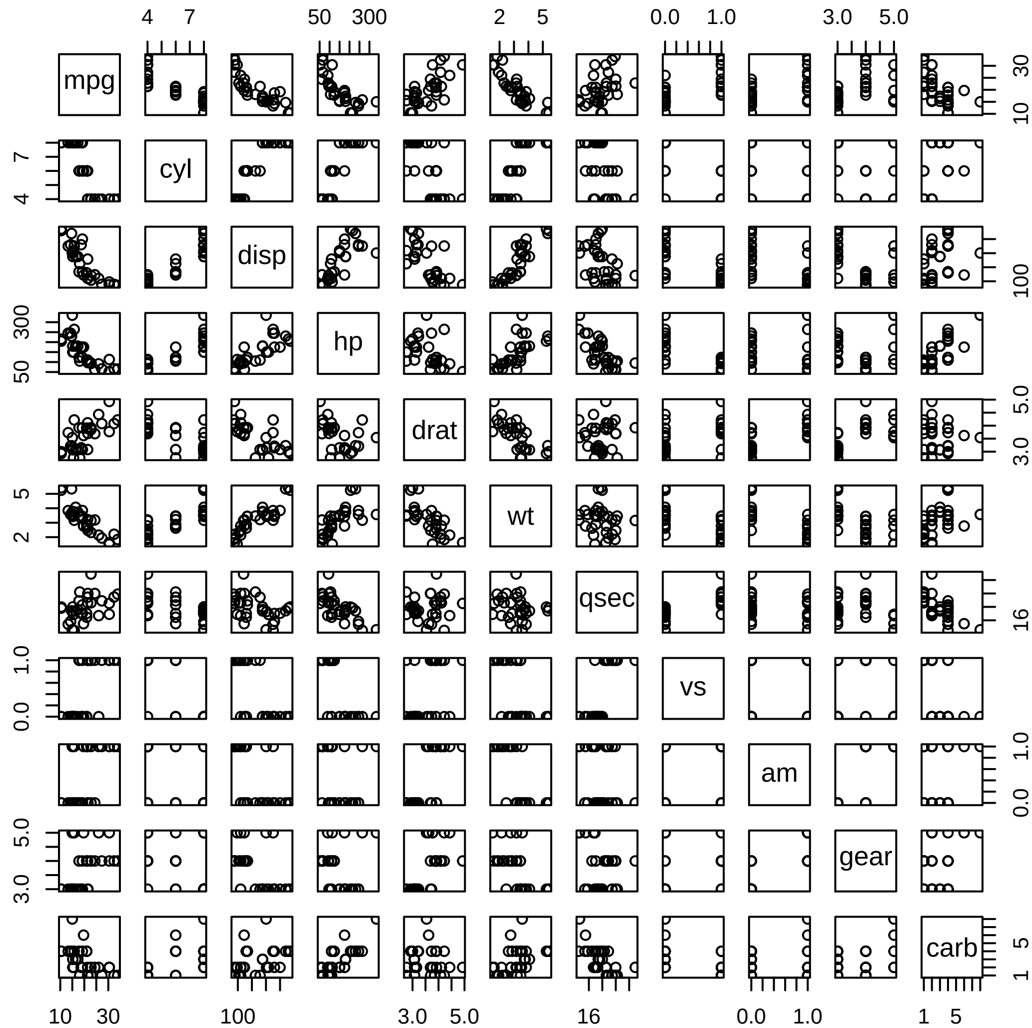

Figure 3: Pairs Plot for mtcars