Modeling MPG by Transmission

Rafi Kurlansik

Monday, September 21, 2015

Summary

This paper aims to determine whether an automatic or manual transmission is better for miles per gallon, and give an estimated quantity for this value. Using the `mtcars` dataset, we will first explore the data and choose our regressors. We will look at two models, and then run some confirmation tests before presenting our conclusions. Please see the appendix for all figures and tables.

Exploratory Data Analysis

The `mtcars` dataset contains 11 variables on 32 different cars from 1973-74. The measured variables are 1)`mpg`, 2)`cyl` number of cylinders, 3) `disp` displacement, 4) `hp` horsepower, 5) `drat` rear axle ratio, 6) `wt` weight(lb/1000), 7) `qsec` 1/4 mile time, 8) `vs` engine orientation(straight or 'V'), 9) `am` transmission (0 = automatic, 1 = manual), 10) `gear` number of forward gears, and 11) `carb` number of carburetors. Table 1 in the appendix displays the original data..

While we ultimately want to know what impact, if any, transmission has on mpg, there are many variables that can have an impact on mpg. The linear model `lm(mpg ~ am, mtcars)` gives us a transmission coefficient of 7.245, which seems to indicate that a manual transmission greatly increases mpg. We also find a positive correlation of 0.599. However, this model is far too simplistic. Many factors outside of transmission can affect mpg - weight, cylinders, possibly others!

Our first task is determine which independent variables appear to have a relationship with the dependent variable, mpg. Then we will perform a simple check for multicollinearity among the independent variables, and based on what we find there we will be able to choose the regressors to build our model.

Table 2 in the appendix is a table with correlation values for each variable in the `mtcars` dataset. Using this table, we can tell that weight (-0.868), cylinders(-0.852), and displacement(-0.848) have the strongest correlations to mpg. Looking for redundancy, weight is strongly correlated with cylinders(0.782) and displacement(0.888). Because weight and displacement are so closely correlated, and weight is more closely correlated to mpg, we will drop displacement from our regression. Cylinders are slightly less correlated, so we will build two models - one including cylinders and one without. Figure 1 shows the relationships between our regressors with scatterplots and correlations.

Regression Models

The stage is set to build two multivariate linear regressions with outcome variable `mpg`. Model 1 contains predictors `am`, `wt`, and `cyl`. Model 2 leaves out cylinders.

```
model1 <- lm(mpg ~ am + wt + cyl, mtcars)  ## First model, including cylinders.
model2 <- lm(mpg ~ am + wt, mtcars) ## Second model, without cylinders.

summary(model1)$coefficients
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 39.4179334  2.6414573 14.9227979 7.424998e-15
## am           0.1764932  1.3044515  0.1353007 8.933421e-01
## wt          -3.1251422  0.9108827 -3.4308942 1.885894e-03
## cyl         -1.5102457  0.4222792 -3.5764148 1.291605e-03
```

```
summary(model2)$coefficients
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

We can interpret the coefficients for `am` (transmission) to mean that in **model 1, we get a -0.024 decrease in mpg for having a manual transmission, while in model 2 we see a 0.177 increase in mpg.** There is a slightly better adjusted r-squared value in model 1: 81% vs. 73%. This tells us that we have explained more of the total variation by including cylinders in our regression. Figure 2 in the appendix illustrates this phenomena, with model 1 having slightly less variation around the regression line than model 2.

95% confidence intervals for both models with $\hat{B_1} \pm 1.96 * SE(\hat{B_1})$ are:

```
sumCoef1 <- summary(model1)$coefficients
sumCoef1[2,1] + c(1, -1) * 1.96 * (sumCoef1[2,2])
## [1]  2.733218 -2.380232
sumCoef2 <- summary(model2)$coefficients
sumCoef2[2,1] + c(1, -1) * 1.96 * (sumCoef2[2,2])
## [1]  3.00585 -3.05308
```

In both models the intervals include 0. Thus, we cannot conclude with 95% confidence that there is a significant relationship between transmission and mpg. The p-values for transmission in these models are also extremely high:

```
sumCoef1[2,4] ## Model 1 transmission estimate p-value
## [1] 0.8933421
sumCoef2[2,4] ## Model 2 transmission estimate p-value
## [1] 0.9879146
```

Were there to be no relationship between mpg and transmission type, we would get these values 89% and 99% of the time. That's extremely poor for trying to establish a meaningful relationship between the two.

Running a `hatvalues` test and residual plot shows that, overall, the model fits the data well. Figure 3 illustrates that there is no discernible pattern in the residuals, and the hat values have no extreme impact on the model.

Conclusion

In conclusion, which kind of transmission one chooses has no significant effect on miles per gallon. At 32 cars, the sample size is somewhat small, but even so, the numbers aren't even close. Based on the two models in this analysis, weight and cylinders have a far greater impact on what kind of mpg you get.

Appendix

Table 1:

```
mtcars
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

Table 2:

```
##             mpg        cyl       disp         hp         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000  0.6587479
## wt   -0.8676594  0.7824958  0.8879799  0.6587479  1.0000000
```

Figure 1:

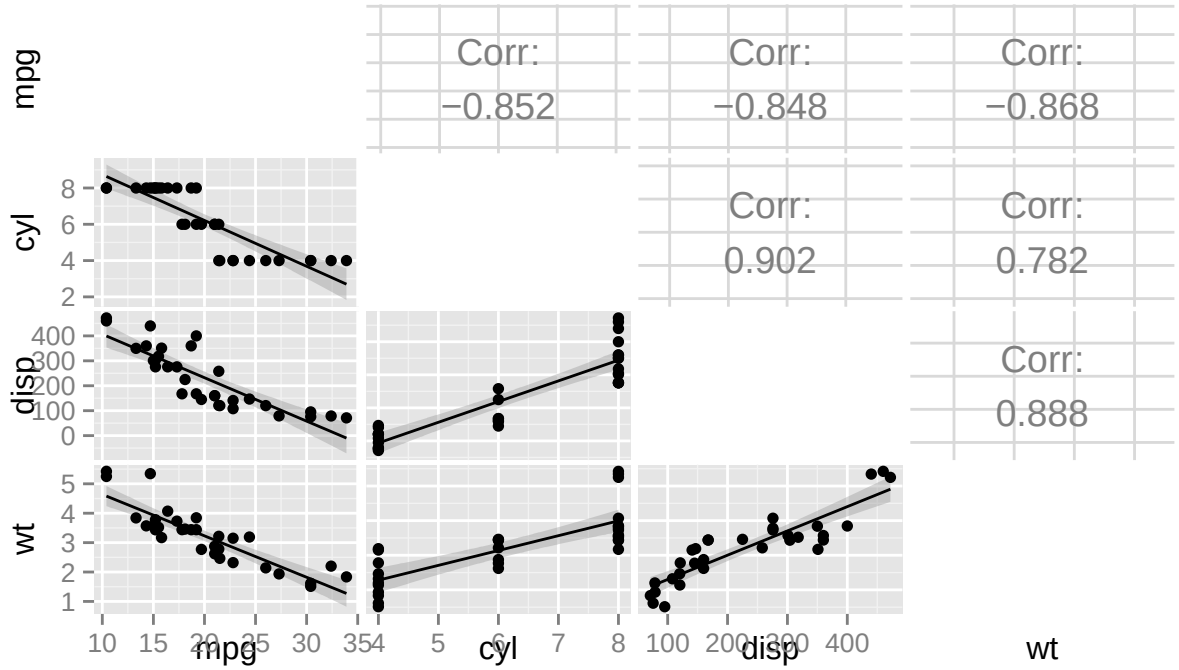


Figure 2:

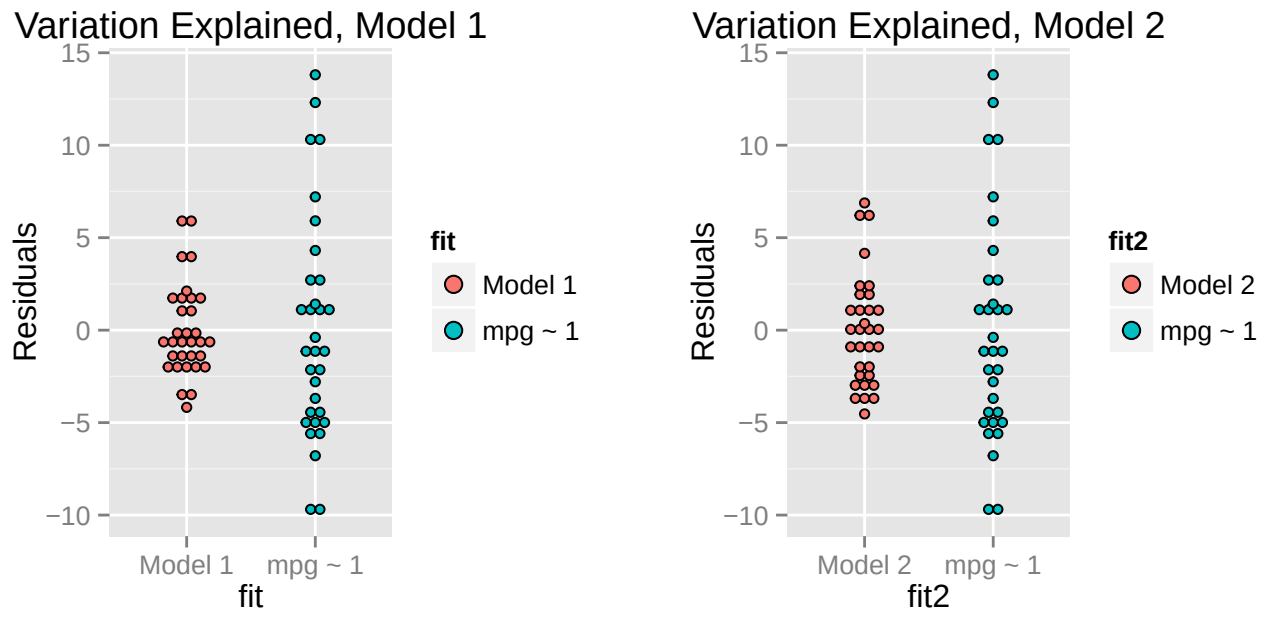
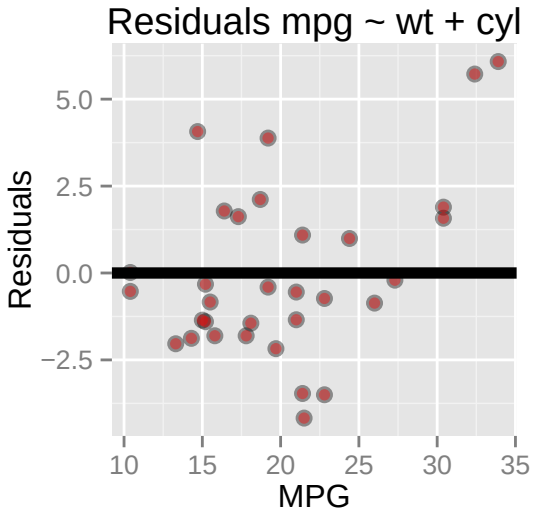
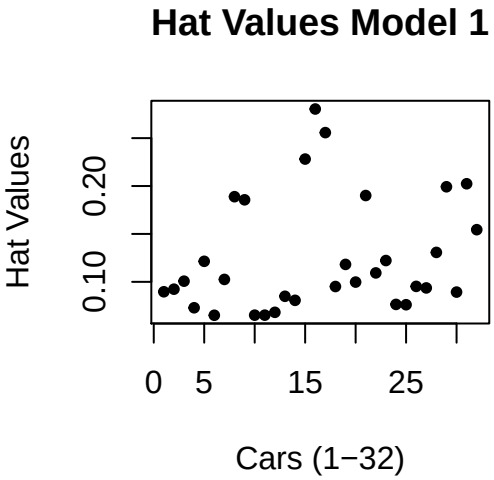


Figure 3:


Residuals mpg ~ wt + cyl
Residuals
5.0
2.5
0.0
−2.5
10
15
20
25
30
35
MPG


Hat Values Model 1
Hat Values
0.10
0.20
0
5
15
25
Cars (1−32)