Evaluating Fuel Economy of Manual vs. Automatic Transmissions using Regression Analysis

Dann Hekman

Saturday, March 07, 2015

Executive Summary

There is a common belief that cars with a manual transmission are more fuel-efficient than automatics. Using 1974 data from *Motor Trends* we can analyze this claim using linear regression analysis. Based on the analysis presented below, manual transmission cars get on average 7.24 more miles per gallon than automatic cars. Although tempered, this main effect remains significant at the 5% confidence level, even when controlling for weight and acceleration of the vehicle.

Research Question

We start our analysis by examining the claim that manual transmission cars have better gas mileage than automatic cars. We can load the data and do some basic analysis in R on this dataset.

```r
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```r
library(ggplot2)
```

Specifically, our variables of interest are mpg and am, which is coded to 0 for automatic transmissions, and 1 for manual transmission. We can regress

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$

where X_i is the *as* variable in the mtcars dataset. Simple Least Squares regression will give the average mpg for automatics ($X = 0$) and β_1 will give the added average miles per gallon of a manual transmission. Looking at just the transmission, we can see that automatic transmission vehicles have an average of 17.15 miles per gallon, and manuals get an additional 7.24 miles per gallon on average (24.39 total).

```r
model <- lm(mpg ~ am, data=mtcars)
summary(model)$coef
```

```
##             Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

Correction for Car Weight

Before jumping to the conclusion that manual transmissions are more fuel efficient, we should try to control for other variables in the dataset.

$$Y_i = \beta_0 + \beta_1 X_i + \sum \beta_j X_i j + \epsilon_i \tag{2}$$

where β_1 is the coefficient for the dummy variable *as* and the sum of β_j and X_j includes all other variables.

```r
model <- lm(mpg ~ ., data=mtcars)
summary(model)$coef
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

Model Selection

Using the `AIC` function in R, we systematically removed variables, one at a time, until we could minimize the AIC output. Below are a subset of the models reviewed:

```r
AIC(lm(mpg ~ wt + cyl + disp + hp + drat + qsec + vs + am + gear + carb,data=mtcars))
```

```
## [1] 163.7098
```

```r
AIC(lm(mpg ~ wt + disp + hp + drat + qsec + am + gear + carb,data=mtcars))
```

```
## [1] 159.7853
```

```r
AIC(lm(mpg ~ wt + disp + hp + drat + qsec + am,data=mtcars))
```

```
## [1] 156.2687
```

```r
AIC(lm(mpg ~ wt + qsec + am,data=mtcars))
```

```
## [1] 154.1194
```

```r
#Removing any more raises the AIC
AIC(lm(mpg ~ qsec + am,data=mtcars))
```

```
## [1] 175.6022
```

```r
AIC(lm(mpg ~ wt + am,data=mtcars))
```

```
## [1] 168.0292
```

```r
AIC(lm(mpg ~ wt + qsec,data=mtcars))
```

```
## [1] 156.7205
```

After controlling for weight and the time to drive one 4th of a mile, manual vs. automatic is still a statistically significant predictor of gas mileage at the 5%, but not at the 1% level, and furthermore, the magnitude is less than the effect of weight once you control for the other two variables. Our final model is

$$Y_i = \beta_0 + \beta_1 wt_i + \beta_2 qsec_i + \beta_3 am_i + \epsilon_i \tag{3}$$

Where wt is the car weight (in 1000 lbs), $qsec$ is the time to drive one quarter mile, and am is a dummy variable where 0=automatic and 1=manual. This is only one of several possible model selection methods, and is notably limited in that it does not account for possible interaction terms or power models.

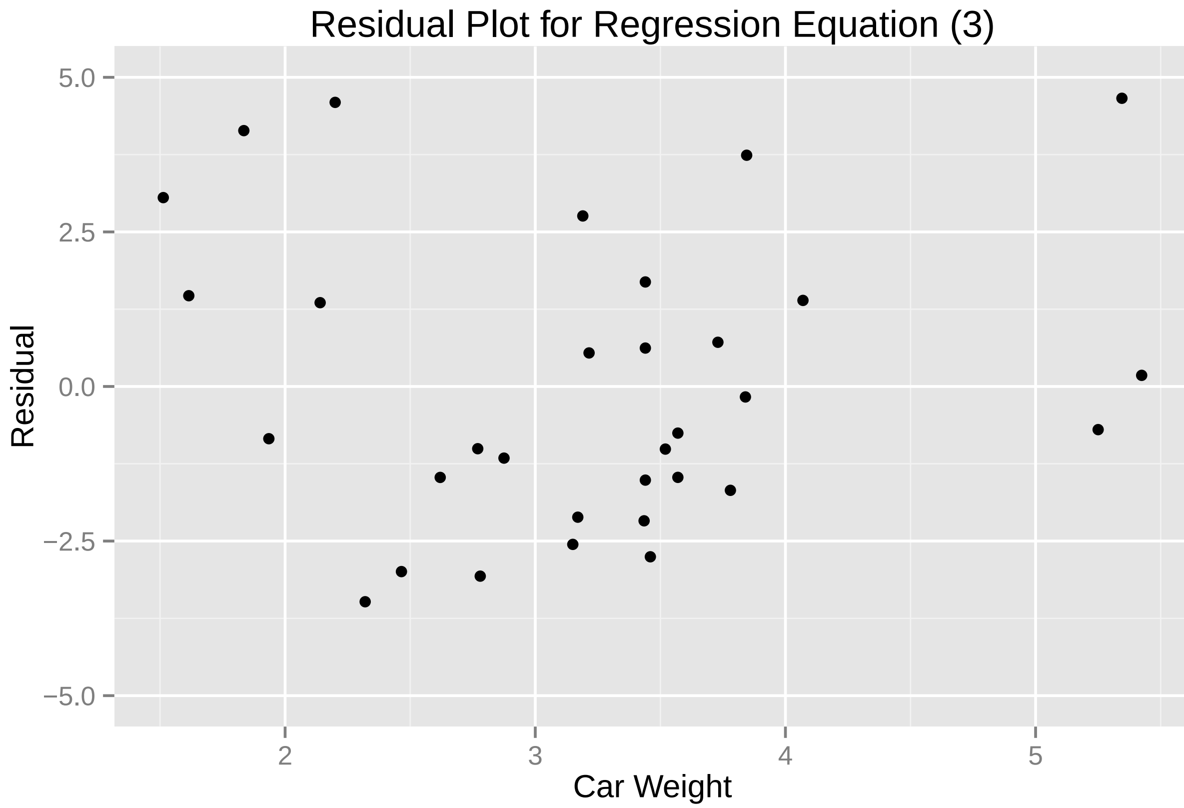
```r
model<-lm(mpg ~ wt + qsec + am,data=mtcars)
summary(model)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am           2.935837  1.4109045  2.080819 4.671551e-02
```

Diagnostics

The validity of a linear regression model requires that the variance of the residuals be constant and that the residuals themselves are not closely correlated to the predictor variables used in the equation. We tested these assumptions using a resitual plot and by regression the residuals against the predictor variables.

```r
qplot(x=mtcars$wt,y=model$resid) + ylim(-5,5) + xlab("Car Weight") + ylab("Residual") + ggtitle("Residu
```




```r
summary(lm(model$resid~wt+qsec+am,data=mtcars))
```

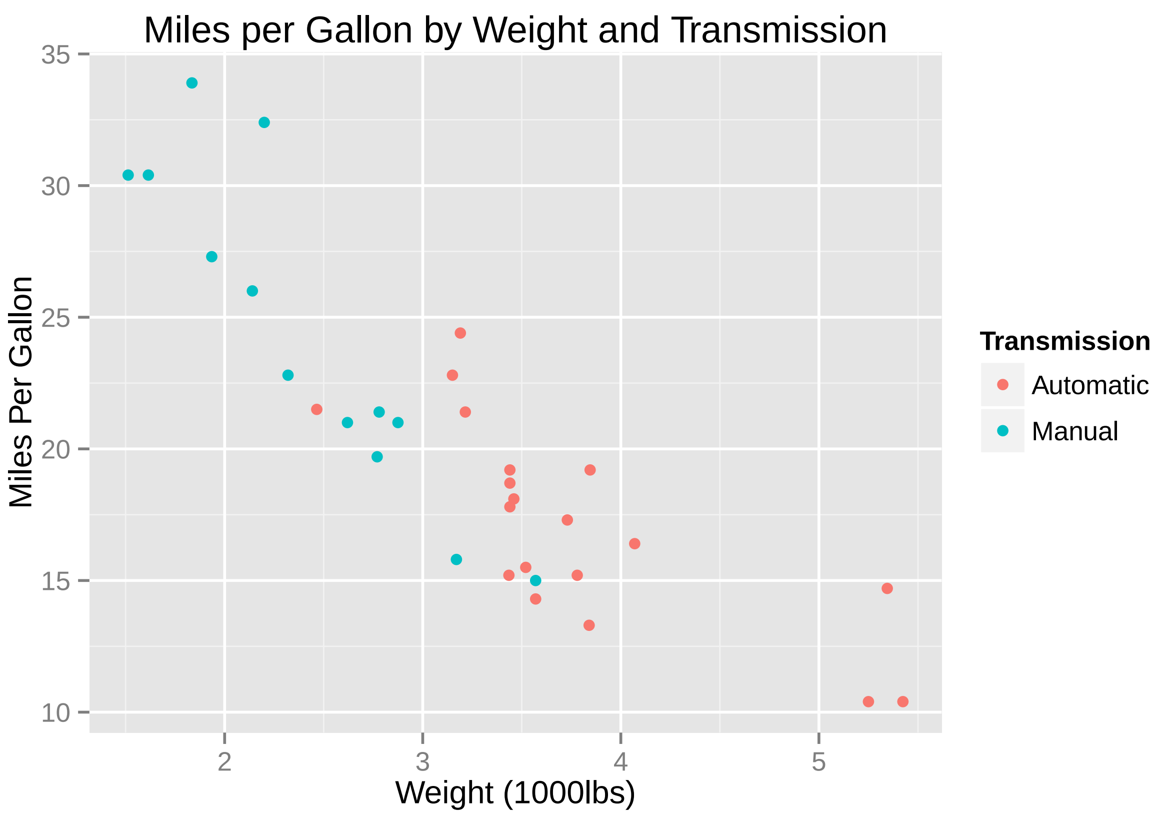
```
##
## Call:
## lm(formula = model$resid ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.561e-15  6.960e+00       0        1
## wt           2.661e-16  7.112e-01       0        1
## qsec         1.422e-16  2.887e-01       0        1
## am           5.096e-16  1.411e+00       0        1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  2.778e-32,  Adjusted R-squared:  -0.1071
## F-statistic: 2.592e-31 on 3 and 28 DF,  p-value: 1
```

Thus we can conclude there is no heterskedasticity and no issues with a non-constant residual variance.

Conclusions

After controlling for other factors and validating the model using the Akaike Information Criterion (AIC), we conclude that manual transmission vehicles do, on average, have a better gas mileage by 2.94 mpg at the 5% confidence level. This relationship between mpg, transmission, and weight is very well illustrated by the following graph.

```
q<-qplot(x=wt,y=mpg,data=mtcars,colour=as.factor(am))
q<-q + scale_color_discrete(name="Transmission",labels=c("Automatic","Manual"))
q<-q + ggtitle("Miles per Gallon by Weight and Transmission")
q<-q + xlab("Weight (1000lbs)") + ylab("Miles Per Gallon")
q
```



So we can see clearly that automatics tend to be both heavier and have a lower gas mileage, but comparibly weighted manual transmission vehicles still tend to have slightly better gas mileage, as seen with equation (3).