Regression project

Deepthi Murthy

June 17, 2015

Executive summary

This brief report summarizes my findings on the impact of transmission type on fuel economy. Using a sample of 32 cars from a 1974 Motor Trends magazine, I find that there is not enough evidence to suggest that the type of transmission has a significant impact on fuel economy. While the mean number of miles per gallon is quite different between the manual and automatic transmission cars, this apparent difference seems to vanish after accoutning for other factors.

The data

Our dataset contains data on 32 different cars. The variables include miles per gallon, number of cylinders, displacement, gross horsepower, rear axle ratio, weight (lb/1000), quarter mile time, engine type, transmission type, number of forward gears, and the number of carburetors.

Highlighting our variables of interest, there are 19 cars with automatic transmissions. They have a mean MPG of 17.15 with a standard deviation of 3.83. There are 13 cars with manual transmissions. They have a mean MPG of 24.39 with a standard deviation of 6.17.

On first glance, then, it appears that cars with a manual transmission are more fuel efficient. However, that may not be the cause; it is possible that the cars with a manual transmission happen to share some other characteristics.

Indeed, many of the possible covariates are highly correlated with each other; see figure 1 in the appendix for pairwise scatter plots. Further, many indicate a significant correlation with the type of transmission. There does not appear to be any cars that would be considered outliers in any of the plots.

Models

The highly correlated covariates make it difficult to identify an accurate model. Figure 2 contains the results of a regression model that includes all covariates in the set. Many of the covariates appear to have small effects other than weight and transmission type. This model estimates an increased fuel economy of 2.52 miles per gallon for manual transmissions, though the result is not statistically significant. Further, including all of the covariates inflates the variance greatly as seen by the variance inflaction factors, so simplifying the model makes sense.

The two covariates that appear to have the largest impact on fuel economy are the type of transmission and the weight of the car. Common sense reasoning backs up these being important factors to consider, so I'll use these variables for the next model.

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

Weight, as expected, has a signficant, negative impact - an estimated drop of 5.35 miles per gallon for every additional 1,000 pounds. However, more interestingly, this model does not provide evidence that the transmission type has a meaningful effect on fuel economy. The estimate of the coefficient is near zero and the p-value is near one.

Two further models that expand the previous model are included - one that includes engine type, and one that includes engine type and the number of cylinders. Both models support the notion that there is not enough evidence to claim that transmission type has a significant effect on fuel economy.

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 30.078722  3.7479768  8.025322 9.709566e-09
## amManual     1.491345  1.4862953  1.003398 3.242621e-01
## wt          -3.784454  0.8981283 -4.213712 2.363336e-04
## vsS          3.615040  1.2760521  2.832988 8.454158e-03
```

```
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 37.4753638  5.2186414  7.1810575 1.008692e-07
## amManual     0.4899110  1.5080744  0.3248587 7.477918e-01
## wt          -3.0946199  0.9270545 -3.3381208 2.470224e-03
## vsS          0.8173533  1.8839753  0.4338450 6.678480e-01
## cyl         -1.2905360  0.6634075 -1.9453141 6.221845e-02
```

Each of the model's residuals can be assumed to normally distributed using the Shapiro test at the 5% level - details are in the appendix. Thus using an ANOVA test to compare these models, the likelihood ratios support including these additional covariates.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + wt
## Model 2: mpg ~ am + wt + vs
## Model 3: mpg ~ am + wt + vs + cyl
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     29 278.32
## 2     28 216.32  1    62.004 8.8239 0.006176 **
## 3     27 189.72  1    26.591 3.7842 0.062218 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The inclusion of additional covariates results in minor changes in the coefficients, but two points seem to hold true. One, the impact of transmission type remains not statistically significant. And secondly, though the details are not included, likelihood ratio tests do not support the addition of too many covariates.

Conclusion

Although the mean fuel economies of manual and automatic tranmissions are significantly different, the notion that there actually is a difference does not hold up to a more thorough analysis. It seems that, accounting for other factors that could influence the fuel economy of a car, that transmission type does not have much of an impact.

Appendix

Figure 1: Pairwise Scatterplots

The red dots represent maual transmissions, and the black dots represent automatic transmissions.

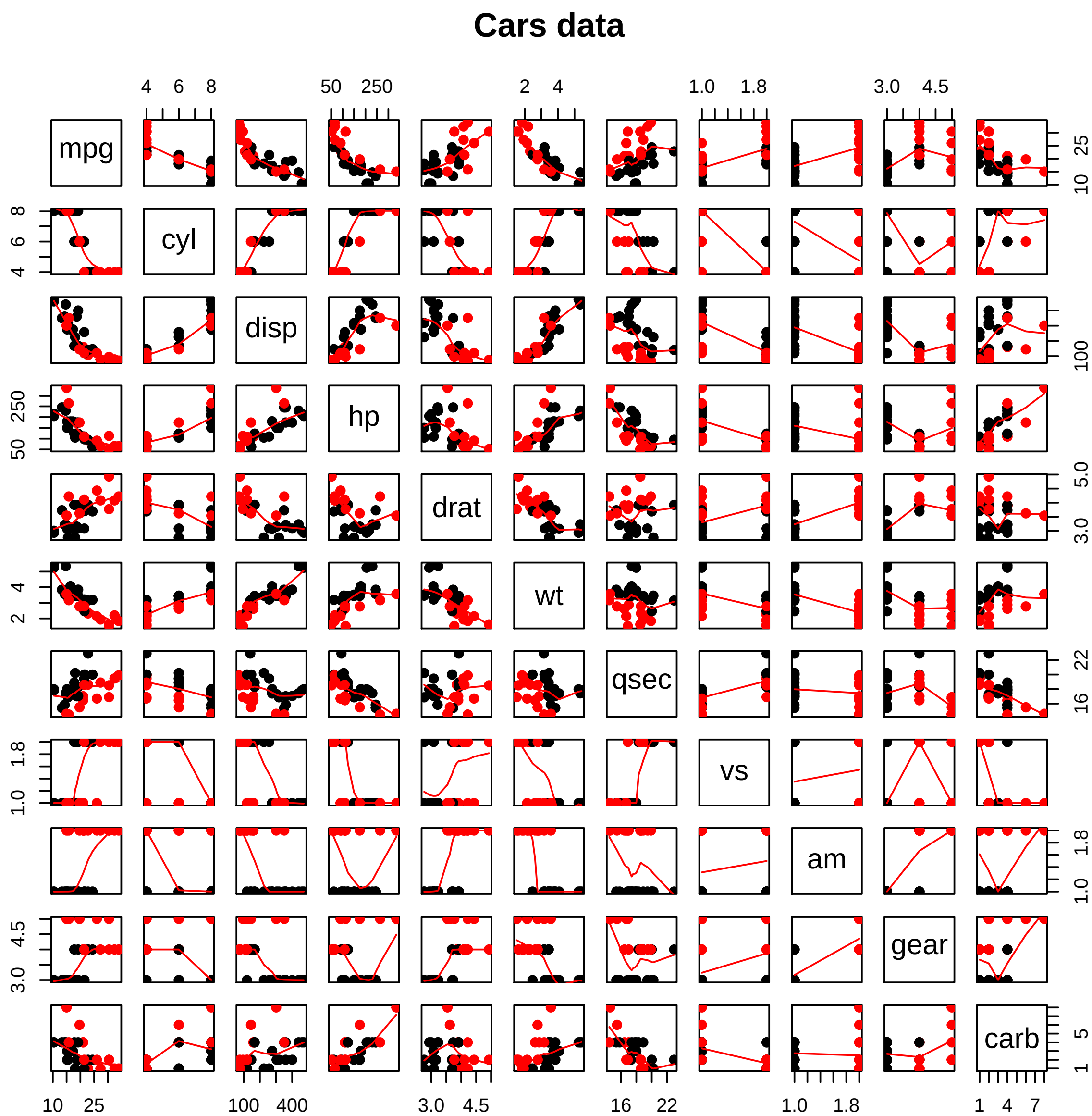


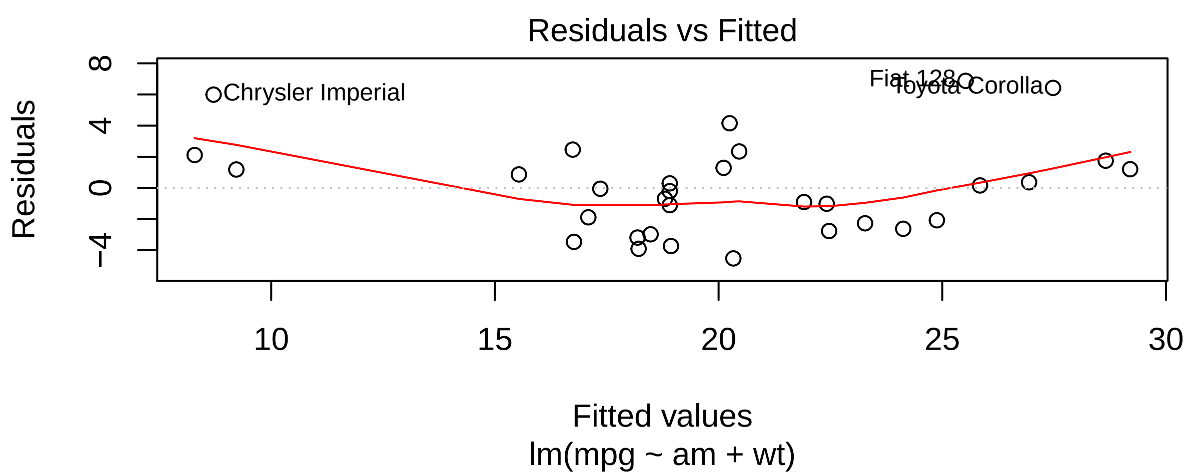
Figure 2: Model 1 - All Covariates

```
##                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
## vsS          0.31776281  2.10450861  0.1509915 0.88142347
## amManual     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
```

```
##       cyl      disp        hp      drat        wt      qsec        vs
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873
##        am      gear      carb
##  4.648487  5.357452  7.908747
```

Figure 3: Model 2 - Weight and Transmission Type



Residual plot

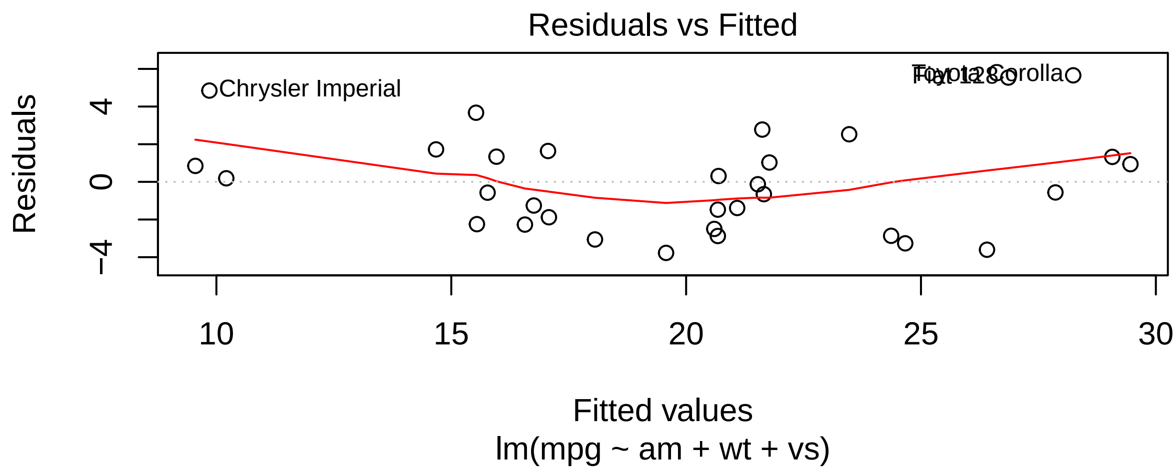
Shapiro test of residual normality

```
##
##  Shapiro-Wilk normality test
##
## data:  model.2$residuals
## W = 0.9448, p-value = 0.1024
```

Variance inflation factors

```
##       am       wt
## 1.921413 1.921413
```

Figure 4: Model 3 - Weight, Transmission, and Engine



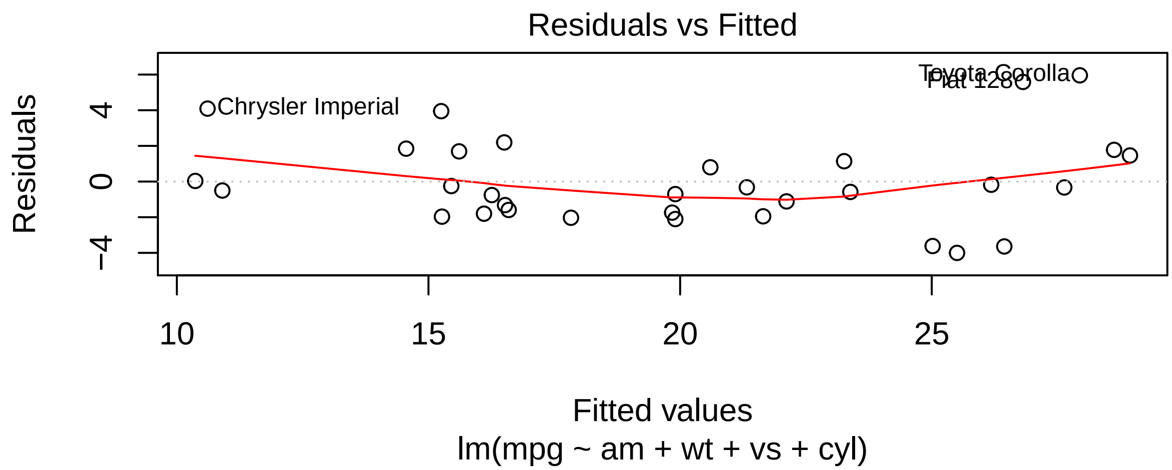
Shapiro test of residual normality

```
##
##  Shapiro-Wilk normality test
##
## data:  model.3$residuals
## W = 0.9484, p-value = 0.1296
```

Variance inflation factors

```
##       am       wt       vs
## 2.207126 3.098791 1.659806
```

Figure 5: Model 4 - Weight, Transmission, Engine, and Cylinders



Shapiro test of residual normality

```
##
##  Shapiro-Wilk normality test
##
## data:  model.4$residuals
## W = 0.9404, p-value = 0.07673
```

Variance inflation factors

```
##       am       wt       vs      cyl
## 2.498233 3.629916 3.977786 6.192809
```