Regression Project: Fuel Economy by Transmission Type

Zain Tejani

Sunday, January 25, 2015

Executive Summary

The mtcars dataset was loaded into R to study the relationship between fuel economy (mpg) and transmission type (automatic or manual, defined by the am column). Using multivariate regression, confounders such as weight (wt), acceleration (qsec), and horsepower (hp) were selected to allow for the best prediction of the effect of transmission type on fuel economy. The adjusted R squared value and residual standard error were studied to quantifiably confirm the best choice of parameters. From studying the model's coefficients, it is clear that manual transmissions are better for fuel economy, though some grouping of the data by weight might intorduce significant uncertainty.

Analysis

Using the following code, the mtcars data was loaded.

```
library(datasets)
data(mtcars)
```

Using the `?mtcars` command in R, one can see that the transmission classification is binary in the "am" column, with a value of 0 representing an automatic gearbox, and 1 representing a manual gearbox.

The column names and classes were explored, to determine which variables could act as confounders to mpg. Some exploratory plots were also made, which can be found in the appendix.

```
apply(mtcars, 2, class)
```

```
##       mpg       cyl      disp        hp      drat        wt      qsec
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##        vs        am      gear      carb
## "numeric" "numeric" "numeric" "numeric"
```

As can be seen by the exploratory plots, manual transmission cars tend to have a higher mpg value. Additionally, it can be seen that the mpg values of cars are heavily correlated to their weight.

The model was initially developed based on knowledge of cars and their workings. As seen in the exploratory plots, heavier cars consume more fuel, so weight was selected as a confounder. Furthermore, cars that accelerate faster tend to burn fuel quicker. Hence, the qsec variable (a drag racing measurement) was selected as aonther confounder.

Beyond this point, confounders were selected based on trial and error, comparing the R-squared values, residual standard error, and Pr(>|t|) values to select the best alternative for the model. After numerous trials, the hp and carb variables were the two best alternatives found.

The coefficient tables for the models can be seen below:

```
summary(lm(mpg~am+wt+qsec+hp, data=mtcars))$coef
```

```
##                Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 17.44019110  9.3188688  1.871492 0.072149342
## am           2.92550394  1.3971471  2.093913 0.045790788
## wt          -3.23809682  0.8898986 -3.638726 0.001141407
## qsec         0.81060254  0.4388703  1.847021 0.075731202
## hp          -0.01764654  0.0141506 -1.247052 0.223087932
```

```
summary(lm(mpg~am+wt+qsec+carb, data=mtcars))$coef
```

```
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 12.8971970  7.4724922  1.725957 0.095783676
## am           3.5113918  1.4874727  2.360643 0.025720752
## wt          -3.4343216  0.8200199 -4.188095 0.000268619
## qsec         1.0191298  0.3377635  3.017288 0.005507044
## carb        -0.4886034  0.4212171 -1.159980 0.256211958
```

It is interesting to note that though the T statistic values for hp and carb are similar, which makes a chocie difficult. However, the performance of the am variable is better in the hp model, exhibiting a smaller standard error and T statistic, with a p value of similar significance to the carb model. The adjusted R squared and residual standard error are compared below for both models.

```
##      Adjusted R Squared Residual Standard Error
## hp            0.8367919                2.434828
## carb          0.8355852                2.443813
```

Though values are highly similar, the hp model has a slight advantage and is the hence the preferred choice.

The am2 factor variable was created to help quantify the difference between the automatic and manual transmisison It can be seen that the model predicts an increase of about 2.925 mpg from automatic to manual.

```
mtcars$am2<-as.factor(mtcars$am)
lm(mpg~am2-1, data=mtcars)$coef
```

```
##     am20     am21
## 17.14737 24.39231
```

However, there is an underlying trend between weight and transmission type, as seen in the exploratory plots.

```
lm(wt~am2-1, data=mtcars)$coef
```

```
##     am20     am21
## 3.768895 2.411000
```

It can be seen that manual cars weigh around 1358 lbs less than automatic cars on average. Though the model accounts for weight as a confounder, the data for automatic and manual transmissions appears grouped, which could lead to considerable uncertainty. In order to reduce this uncertainty, the interaction between the am2 factor variable with the wt variable was studied.

```
lm(mpg~am2*wt+qsec+hp, data=mtcars)
```

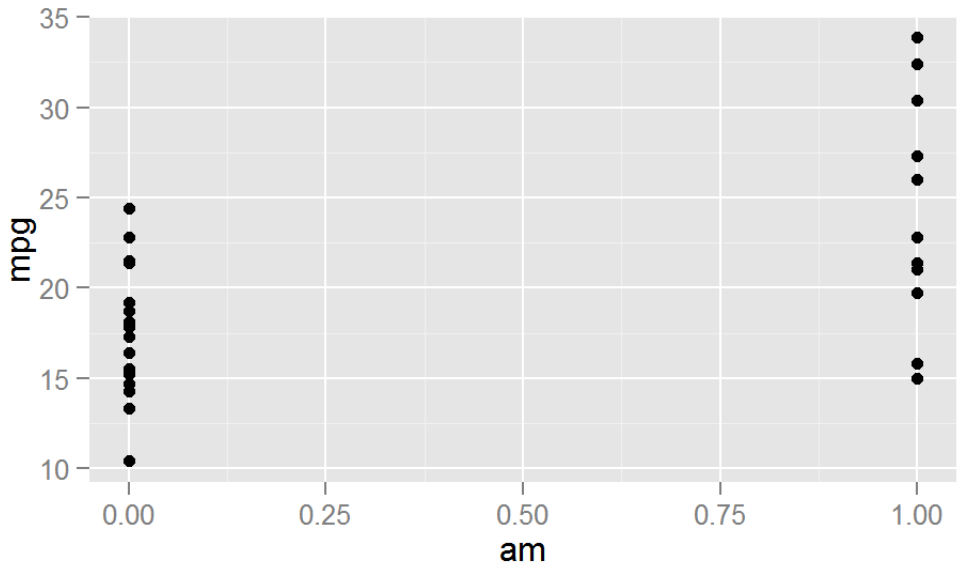
```
##
## Call:
## lm(formula = mpg ~ am2 * wt + qsec + hp, data = mtcars)
##
## Coefficients:
## (Intercept)         am21           wt         qsec           hp
##   10.230790    13.957285    -2.903080     0.992235    -0.001148
##     am21:wt
##   -4.096233
```

It can be seen that a unit change in the weight variable reduces mpg values far more in manual cars than automatic cars, which contradicts the original model. Due to the skewedness of the data by weight, it is difficult to determine the definite effect of the transmission type

Appendix

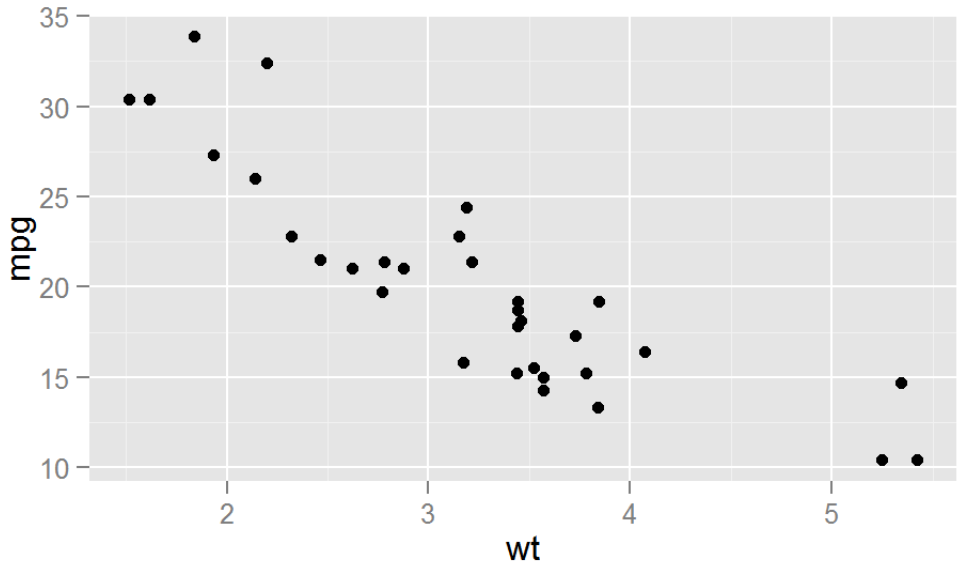
Exploratory plot 1: mpg vs transmission type (am, where 0 is automatic and 1 is manual)

```
library(ggplot2)
qplot(am,mpg, data=mtcars)
```



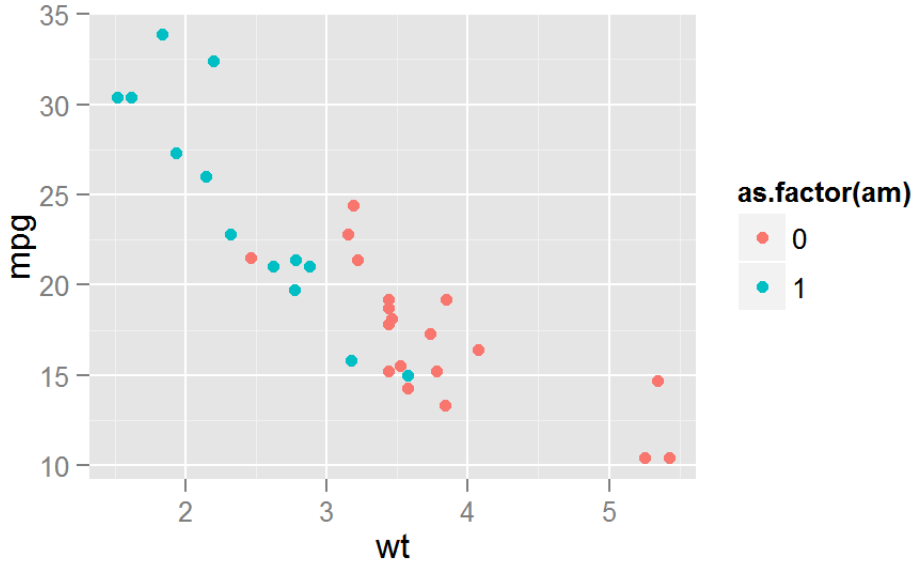
Exploratory plot 2: mpg vs weight

```
qplot(wt,mpg, data=mtcars)
```

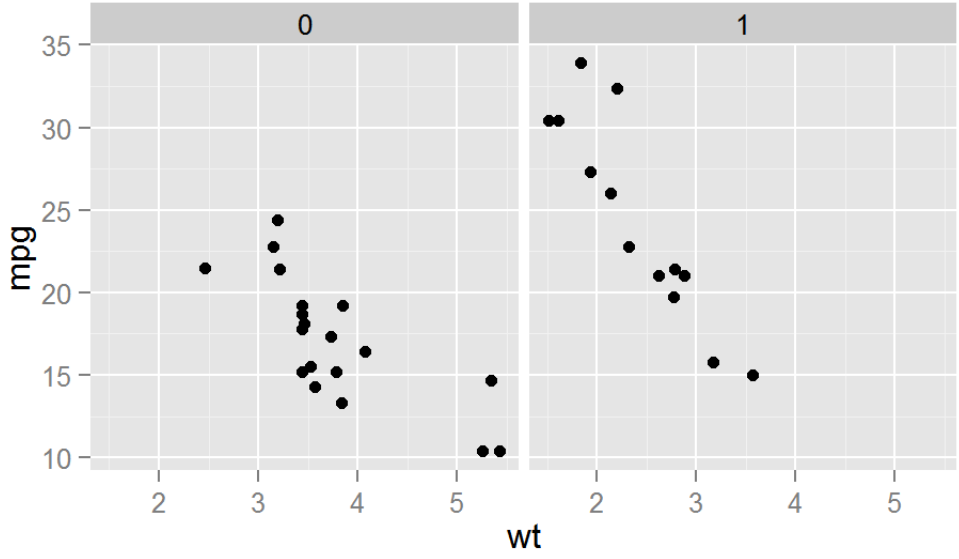


Exploratory plot 3 and 4: mpg vs weight, accounting for transmission type

```
qplot(wt,mpg, data=mtcars, color=as.factor(am))
```

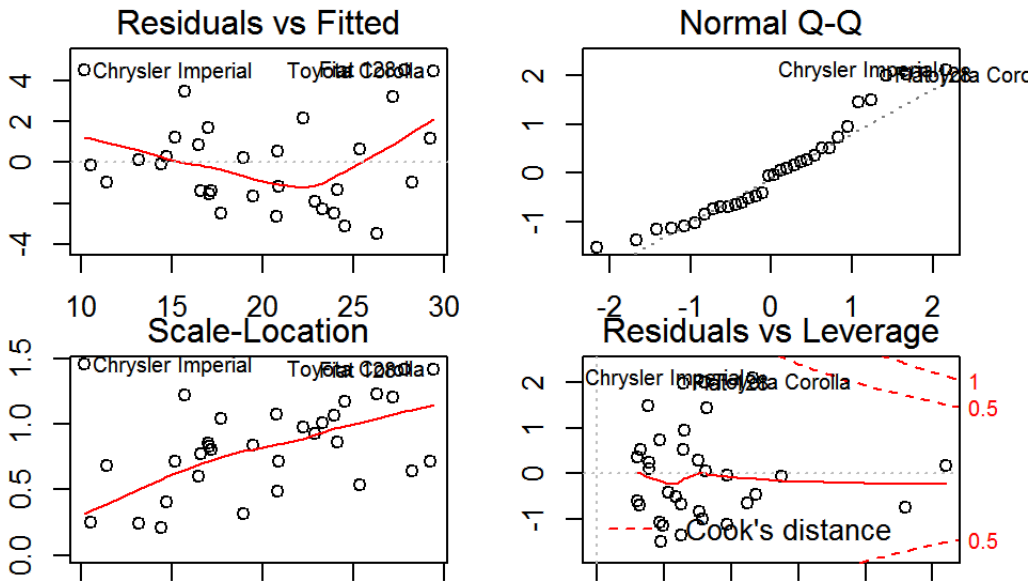



```
qplot(wt,mpg, data=mtcars, facets=.~am)
```



Residual plot 1: hp model

```
par(mfrow=c(2,2), mar=c(1,2,2,2))
plot(lm(mpg~am+wt+qsec+hp, data=mtcars))
```



Residual Plot 2: carb model

```
par(mfrow=c(2,2),mar=c(1,2,2,2))
plot(lm(mpg~am+wt+qsec+carb, data=mtcars))
```