Peer Assignment 1 - Tranny Regression

Eric Smith

December 22, 2015

Abstract

Here we sought to understand the dataset mtcars and how the factor of transmission(tranny) effects MPG. Included in this dataset are numerous measurements for each vehicle. Examining three different models we identify that tranny has a small but significant effect on fuel economy. Using a simple linear model where we have only two non-interacting factors (tranny and weight) we find that the tranny accounts for only -0.02362 in MPG. In other words, having a manual tranny will result in a loss of 0.02 MPG in fuel economy. If we use a more complicated linear model where we allow an interactive term between the two factors we identify that tranny does indeed have a greater impact on MPG. However, the effect is dependent upon weight and therefore saying it is more or less efficient no longer makes sense. We identified that weight account for most of the variation in MPG.

Results

MPGvsTranny

The exploratory analysis shown at the end of this report reveals some interesting features of the MTCars database in regard to the automatic or manual tranny First and foremost we identify that there are a number of parameters beyond mpg that have stark differences between the groups of automatic (auto) and manual (man) that include; cyl, disp, hp, drat, wt, gear, and carb. Based on these finding we cannot attribute the deceased average mpg in the automatic group to that factor. Rather, the decreased mpg could be a result of a confounding variable. For example, weight is greater in the auto group and we know weight to impact mpg greatly.

```
fit1 <- lm(mpg~am, data = MTCars)

summary(fit1)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

Of course, this linear model (lm) is really simple and is not all the useful as we can see from the code below.

```
library(ggplot2)
mean(MTCars[MTCars$am == 0,1])
```

```
## [1] 17.14737
```

```
mean(MTCars[MTCars$am == 1,1])
```

```
## [1] 24.39231
```

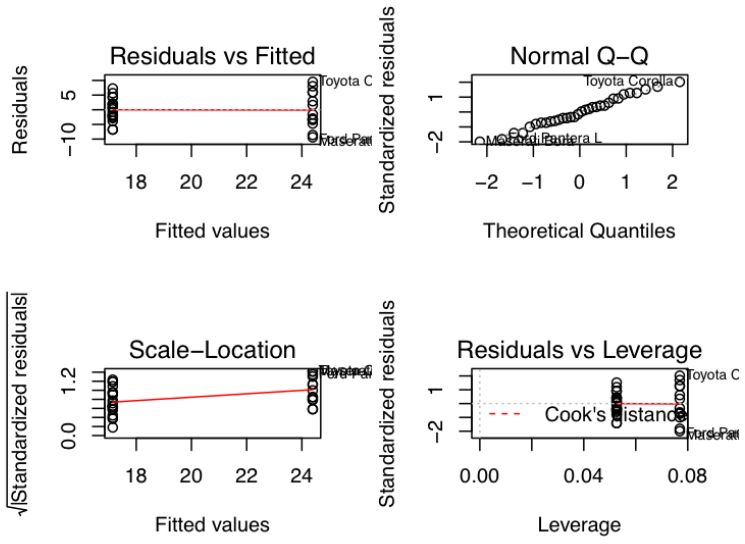
```
mean(MTCars[MTCars$am == 0,1])-mean(MTCars[MTCars$am == 1,1])
```

```
## [1] -7.244939
```

This model simply gives us the average of the two groups MPG in regards to tranny.

The lm with just these two factors, mpg and tranny, calculates that the factor of manual will increase fuel economy by 7.245 MPG. The adjusted R-squared value is just 0.3385. We have large amounts of resudules as can be seen in the following plots.

```
par(mfrow = c(2,2))
plot(fit1)
```

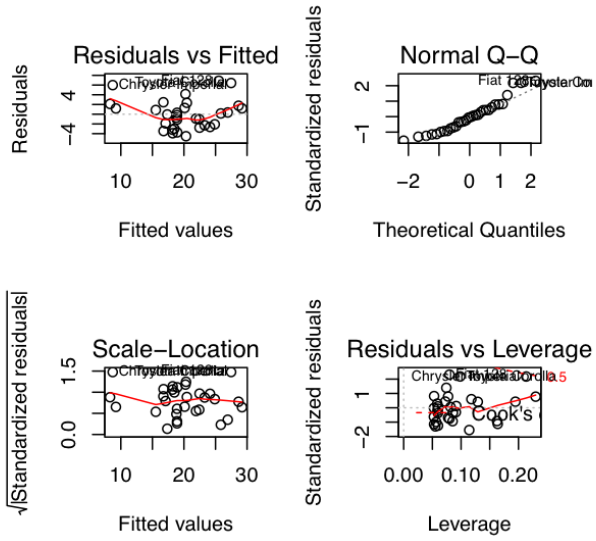


MPGvsTranny+WT

Let's add additional data to see if there are any factors that could influence this difference.

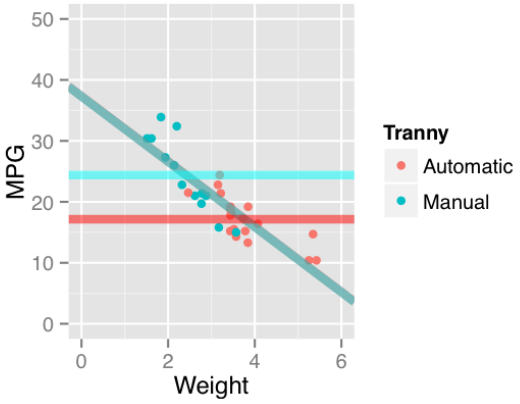
From the ggpairs there are a number of factors we could include. However, it appears that most, if not all, of the correlation seen in the data could be explained by weight.

```
fit2 <- lm(mpg~am + wt, data = MTCars)
par(mfrow = c(2,2))
plot(fit2)
```




```
summary(fit2)$coef
```

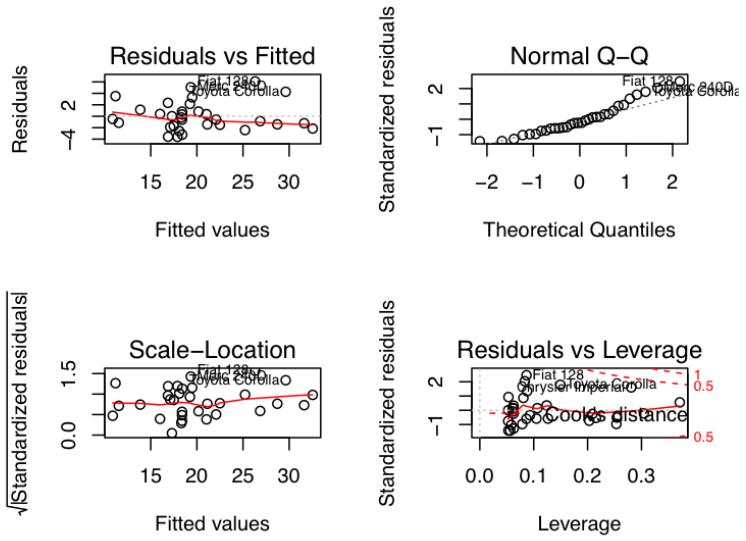
```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am1         -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```



The adjusted R-squared value of 0.7358 is not bad. The graph above shows that both regression lines, one for Automatic and one for Manual lie on top of each other. There is in effect no effect of the factor tranny on the MPG of the car. Let's include an interaction factor to see if we can improve the fit.

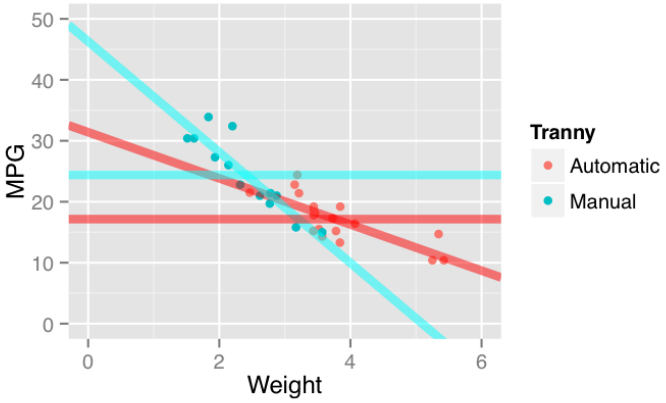
MGPvsWT*Tranny

```
fit3 <- lm(mpg~am * wt, data = MTCars)
par(mfrow = c(2,2))
plot(fit3)
```



```
summary(fit3)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 31.416055  3.0201093 10.402291 4.001043e-11
## am1         14.878423  4.2640422  3.489277 1.621034e-03
## wt          -3.785908  0.7856478 -4.818836 4.551182e-05
## am1:wt      -5.298360  1.4446993 -3.667449 1.017148e-03
```



Finally we can test to see if the inclusion of additional factors improves the model.

```
anova(fit1, fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am * wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 65.913 7.717e-09 ***
## 3     28 188.01  1     90.31 13.450  0.001017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This tells us that even after adding the interaction factor in model 3 we significantly improve the model. In other words there is a .1% chance we would see this difference in models if interaction between weight and transmission did not exist.

Conclusion

In total we find that in regard to MPG tranny has a small but significant effect.

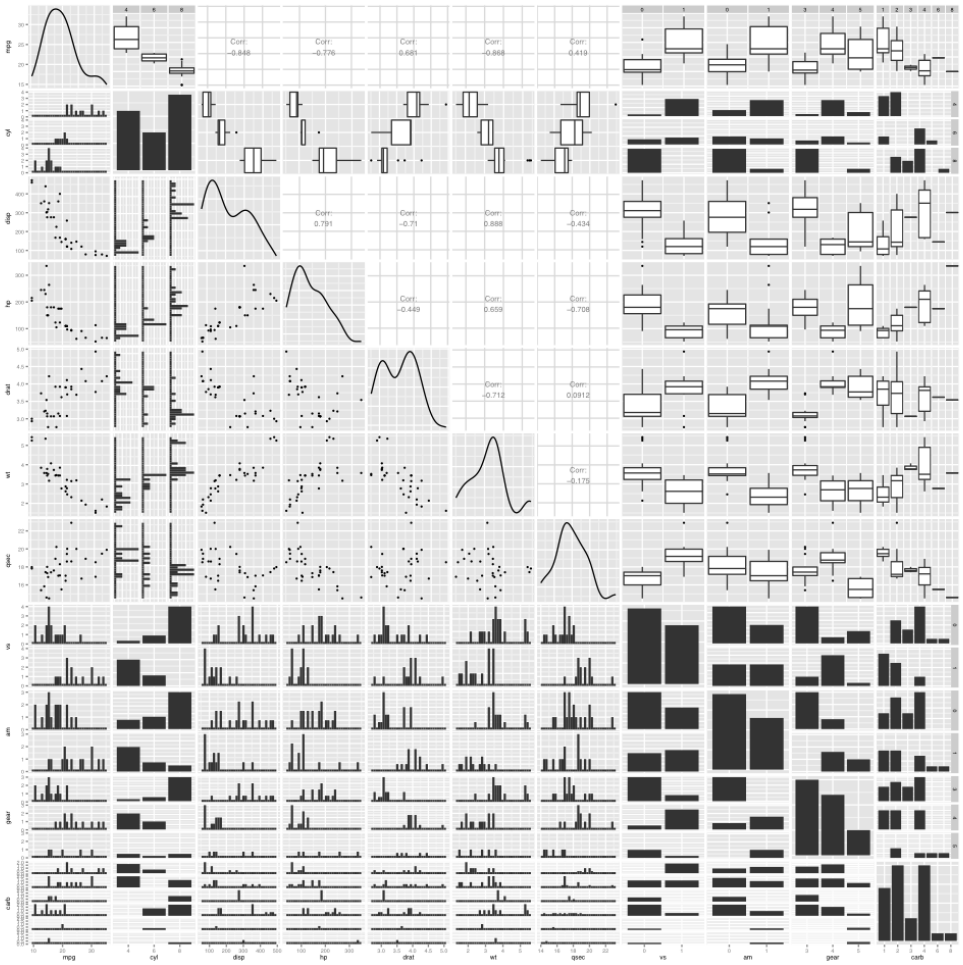
Extra figures

Initial examinations are done to load/clean the data and then understand the data.

```
library(datasets)
library(ggplot2)
library(GGally)
library(dplyr)
data("mtcars")

MTCars <- mtcars

MTCars <- transform(MTCars, carb = factor(carb))
MTCars <- transform(MTCars, gear = factor(gear))
MTCars <- transform(MTCars, am = factor(am))
MTCars <- transform(MTCars, vs = factor(vs))
MTCars <- transform(MTCars, cyl = factor(cyl))
```



This plot was used in the initial exploration of the data.