Influence of Transmission Type on Car Fuel Efficiency

Alex Van Russelt

20 December 2017

Summary

This document addresses the task set by the Coursera Regression Models project, which was to investigate the difference in fuel efficiency between automatic and manual cars.

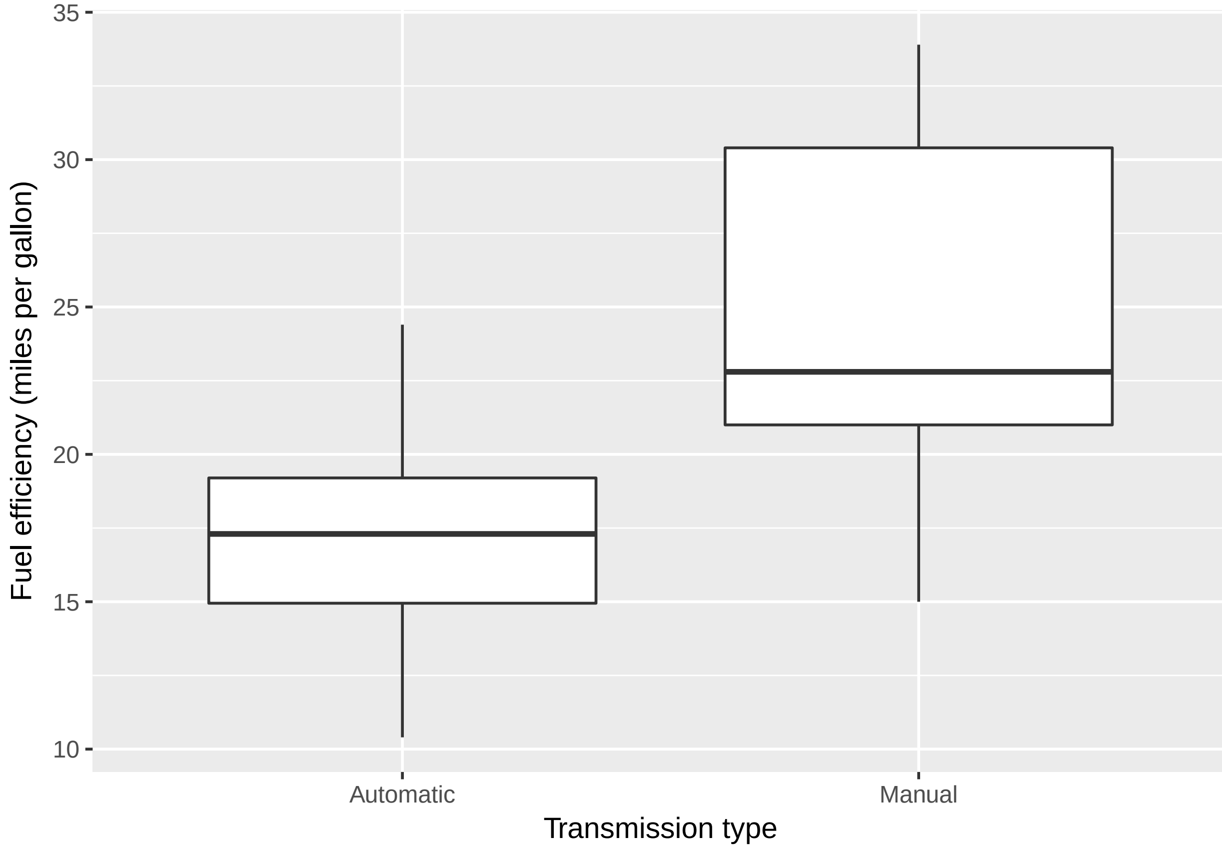
By applying a linear model that took into account the weight of the car, the type of transmission, and the interaction between those two parameters, it was found that cars with a weight less than 2800 lbs were more fuel efficient if they had manual transmissions rather than automatic.

Exploration

```r
library(ggplot2)

data(mtcars)
mtcars$am <- factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

```r
ggplot(mtcars, aes(am, mpg)) + geom_boxplot() +
  labs(x='Transmission type', y='Fuel efficiency (miles per gallon)')
```



On average, manual appear to be significantly more fuel efficient that automatics. However, this could be confounded by other parameters, such as vehicle weight or engine displacement.

Model

As shown in the lecture, a large amount of the variance in mpg can be explained by weight alone, so it would make sense to include this variable in the linear model:

```
fit1 <- lm(mpg ~ am + wt, data=mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
## amManual    -0.02362    1.54565  -0.015    0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

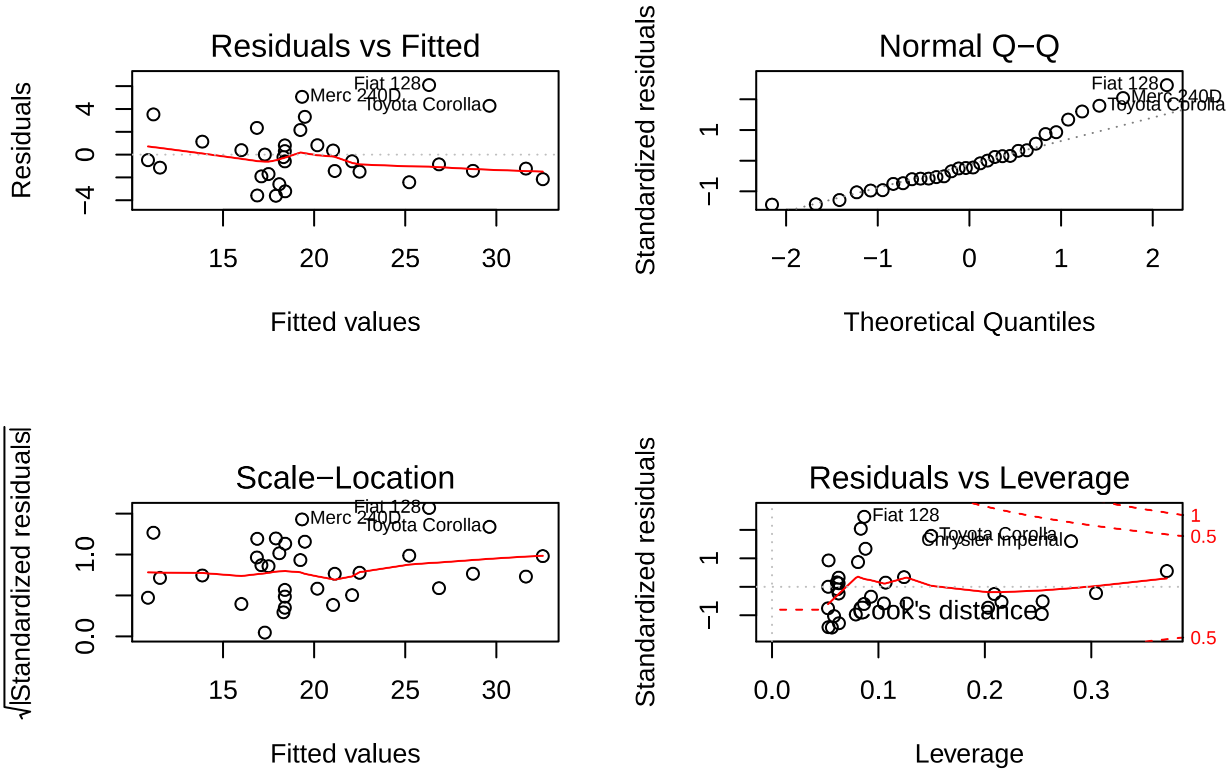
However, the coefficient for `am1` has a *p*-value close to 1, so this particular model may be rejected. Instead, we could consider the interaction between `am` and `wt`:

```
fit2 <- lm(mpg ~ factor(am) * wt, data=mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) * wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6004 -1.5446 -0.5325  0.9012  6.0909
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          31.4161     3.0201  10.402 4.00e-11 ***
## factor(am)Manual     14.8784     4.2640   3.489  0.00162 **
## wt                   -3.7859     0.7856  -4.819 4.55e-05 ***
## factor(am)Manual:wt  -5.2984     1.4447  -3.667  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.591 on 28 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8151
## F-statistic: 46.57 on 3 and 28 DF,  p-value: 5.209e-11
```

This appears to result in a better model, since all coefficient *p*-values are significantly less than the typical threshold of 0.05, the residual standard error has reduced, and the R^2 has increased.

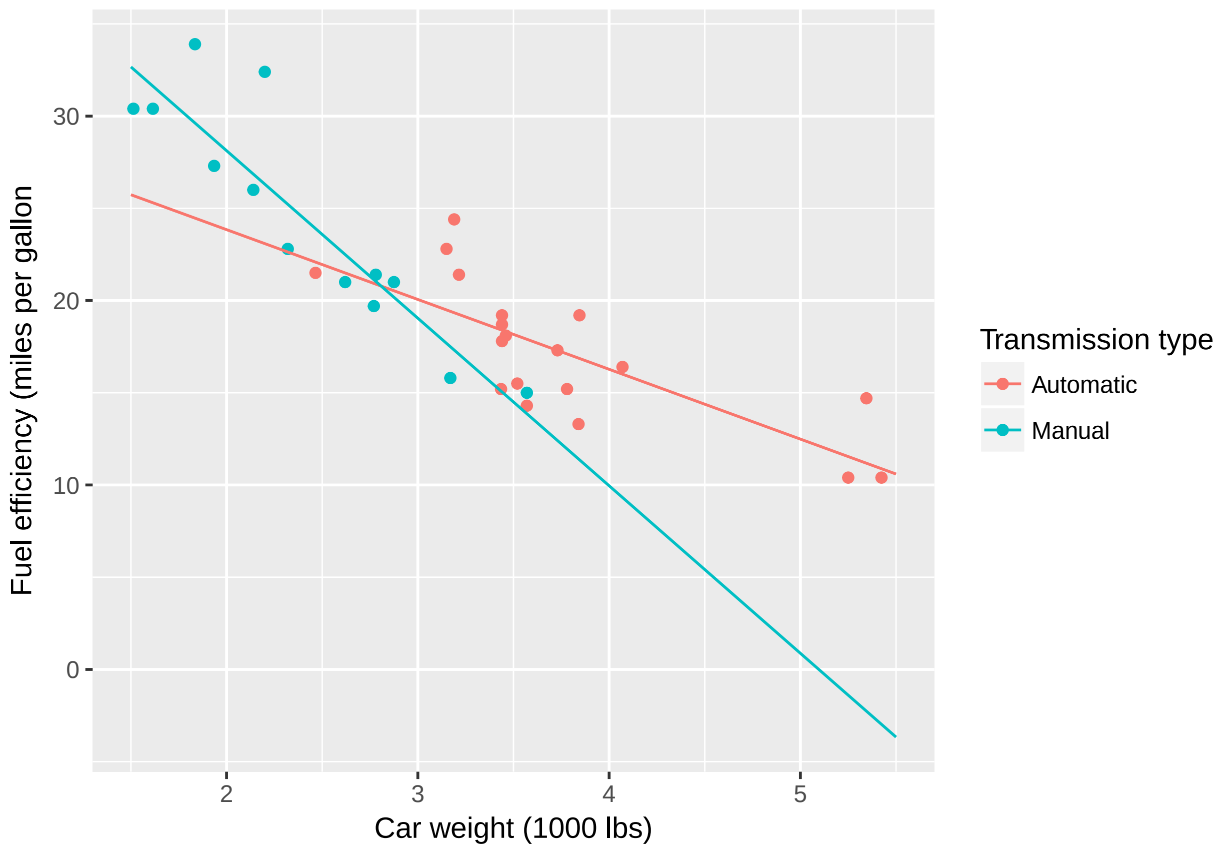
```
par(mfrow=c(2, 2))
plot(fit2)
```



The residual plots do not show any anomalous trends, although the upper tail of the distribution of residuals does not appear to be normal.

Let's visualise the model:

```
am <- factor(c(rep("Automatic", 5), rep("Manual", 5)))
wt <- rep(seq(1.5, 5.5, length=5), 2)
df <- data.frame(am, wt)
df$mpg <- predict(fit2, df)
ggplot(mtcars, aes(x=wt, y=mpg, col=am)) + geom_point() + geom_line(data=df) +
  labs(x='Car weight (1000 lbs)', y='Fuel efficiency (miles per gallon)',
       color='Transmission type')
```



There appears to be a critical car weight above which automatic vehicles have higher fuel efficiency that manual vehicles. The critical value can be calculated by dividing the `factor(am)Manual` coefficient by the `factor(am)Manual:wt` coefficient to show that the critical weight is approximately 2800 lbs:

```
14.878 / 5.2984 * 1000
```

```
## [1] 2808.018
```

The difference in fuel efficiencies can be calculated by looking at the gradients for each transmission, which are different by approximately 1.5 mpg:

```
5.2984 - 3.7859
```

```
## [1] 1.5125
```

Conclusions

We can conclude with a confidence level $p < 0.05$ that for the cars present in this dataset:

- For every 1000 lbs *below* 2800 lbs a car's weight was, a manual transmission was 1.5 mpg *more* fuel efficient than an equivalent automatic.
- For every 1000 lbs *above* 2800 lbs a car's weight was, a manual transmission was 1.5 mpg *less* fuel efficient than an equivalent automatic.

In addition, the 2 sigma prediction interval associated with the model described above is 5.2 mpg.