Regression Models Course Project

Greg D

21 September 2018

Effect of Transmission Choice on Fuel Efficiency

Fuel efficiency is an important metric to consider when purchasing a car, as fuel consumption has both economic and environmental impacts.

This study looks at the effect of transmission choice (manual or automatic) on fuel consumption, using the mtcars dataset from 'Motor Trend' magazine 1973-1974.

Executive Summary

A linear model was fit to the mtcars dataset, to study the effect of transmission choice on fuel efficiency. It was found that the effect was significant, with automatic cars consuming approximately 2.9 mpg more than manual cars.

Overview of dataset

Firstly lets look at a summary of the dataset

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

From the documentation, the 'am' variable is the transmission type, with 0 indicating an automatic, and a 1 indicating a manual.

Let's do a very basic comparison of the two types, but simply looking at the average fuel consumption of cars with each transmission type.

```
mtcars$transmission <- as.factor(mtcars$am)
levels(mtcars$transmission)<- c("auto", "manual")

mean(mtcars$mpg[mtcars$transmission=="auto"])
```

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

```
mean(mtcars$mpg[mtcars$transmission=="manual"])
```

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

It can be seen that, of the models in the dataset, manual cars had an average fuel consumption of 17.1 mpg, compared to 24.4 mpg for automatic vehicles.

This is a bit simplistic though. It might be that automatic transmissions tend to be fitted to larger cars, or cars with larger engines.

Let's fit a linear model and see which factors have the greatest effect on fuel consumption, and try and see the effect of transmission type on its own.

```
fit <- lm(mpg ~ ., mtcars)

summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients: (1 not defined because of singularities)
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        12.30337   18.71788   0.657   0.5181
## cyl                -0.11144    1.04502  -0.107   0.9161
## disp                0.01334    0.01786   0.747   0.4635
## hp                 -0.02148    0.02177  -0.987   0.3350
## drat                0.78711    1.63537   0.481   0.6353
## wt                 -3.71530    1.89441  -1.961   0.0633 .
## qsec                0.82104    0.73084   1.123   0.2739
## vs                  0.31776    2.10451   0.151   0.8814
## am                  2.52023    2.05665   1.225   0.2340
## gear                0.65541    1.49326   0.439   0.6652
## carb               -0.19942    0.82875  -0.241   0.8122
## transmissionmanual       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Let's try dropping a few of the features that don't correlate with mpg (coefficients near 0) and see if we can improve our fit at all.

```
fit2 <- lm(mpg ~ cyl+wt+qsec+am+drat+gear, mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + wt + qsec + am + drat + gear, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5855 -1.5440 -0.7428  1.1166  4.6212
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.3060    17.8380   0.858 0.399014
## cyl          -0.3705     0.8436  -0.439 0.664321
## wt           -3.5659     0.9543  -3.737 0.000971 ***
## qsec          0.9599     0.5885   1.631 0.115388
## am            2.6199     1.9512   1.343 0.191432
## drat          0.6084     1.5544   0.391 0.698789
## gear         -0.4979     1.1601  -0.429 0.671466
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.577 on 25 degrees of freedom
## Multiple R-squared:  0.8526, Adjusted R-squared:  0.8172
## F-statistic:  24.1 on 6 and 25 DF,  p-value: 2.943e-09
```

Lets see if simplifying things further improves our results:

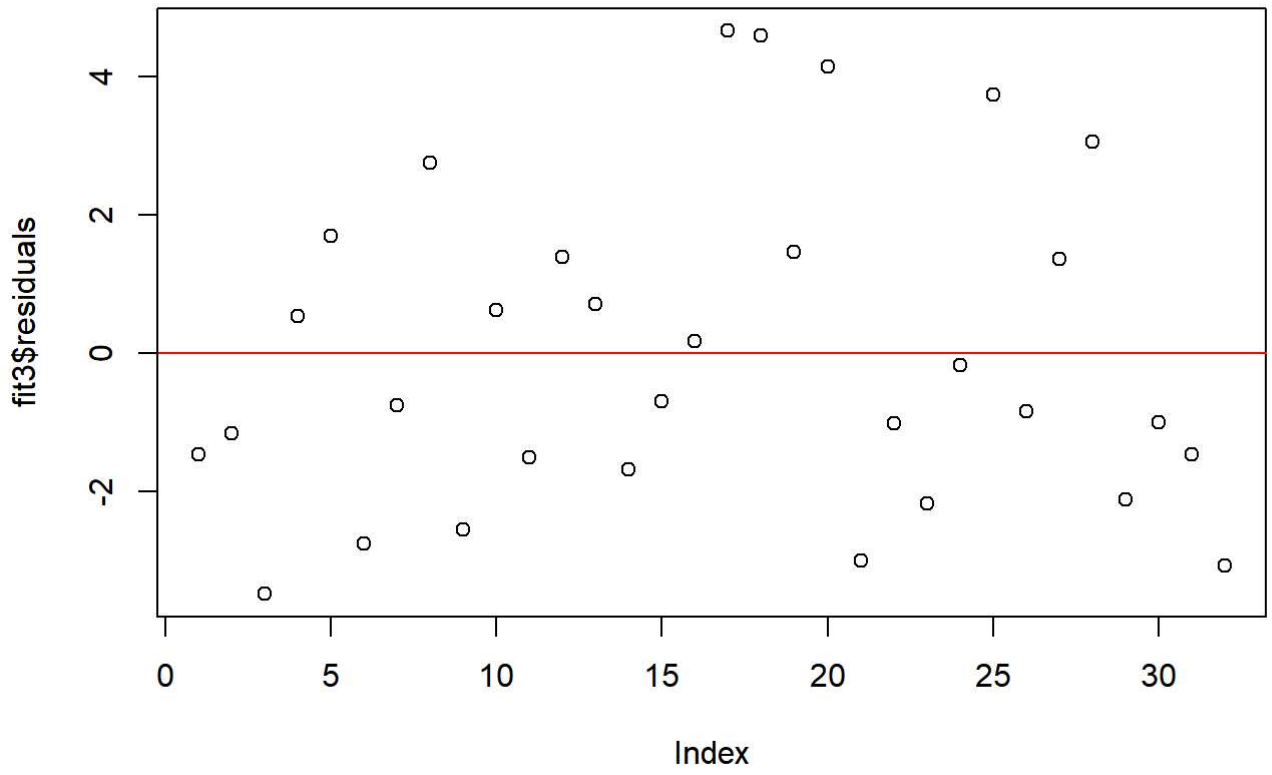
```
fit3 <- lm(mpg ~ wt + qsec + am, mtcars)
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ 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)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

This model has the greatest adjusted R-squared value, and the lowest standard error for the transmission factor.

Let's plot the residuals of this model to see if there are any discernable patterns:

```
plot(fit3$residuals)
abline(h=0, col='red')
```



Residuals are evenly distributed around the zero line, with no obvious pattern. It is unlikely there is significant improvement available with a linear model.

Conclusion

It was found that, correcting for the other significant factors, transmission does have a significant impact on fuel efficiency: an automatic transmission adds approximately 2.9 mpg (standard error 1.4 mpg) to the fuel consumption of a vehicle.