# Motor Trend Fuel Economy by Transmission Analysis.

Iain Russell

20/07/2020

## Executive Summary

The key questions for this project are, firstly whether automatic or manual transmissions are better for fuel economy (mpg), and secondly the quantification of mpg difference between automatic and manual transmissions.

The null hypothesis is that there is no significant difference in mpg between automatic and manual transmissions. Evidence will be presented to either accept or reject this initial hypothesis.

It is reasonable to assume that mpg depends on other variable such as engine displacement and vehicle weight, in addition to the type of transmission in the vehicle. Therefore 3 regression models are used, each with mpg as the outcome; firstly transmission type alone, secondly including engine displacement, and thirdly including vehicle weight.

Considering the general situation of transmission type and nothing else, the mpg difference between Automatics and Manuals is significant at the 95% confidence level ($p < 0.05$). The situation is complicated however by consideration of other factors such as engine displacement and vehicle weight, and when these variables are held constant (controlled for), the difference in mpg between Automatics and Manuals is no longer significant ($p > 0.05$).

With mpg as outcome and transmission category as predictor the residual distribution is approximately normal (Shapiro-Wilks test, $p = 0.8573$), and the residual plot indicates no particular pattern which might invalidate regression assumptions.

When considering transmission type alone, the median mpg is 17.1 for Automatics and 24.4 for Manuals, which can be seen in the boxplot in the appendix. The 95% confidence interval for these 2 estimates overlap however, because mpg fuel efficiency is influenced by other variables beside transmission type.

## Dependencies

Libraries; dplyr for dataframe manipulation, broom for neat regression results display, ggplot2 for data plots.

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

The 'mtcars' dataset. Set new factor variable 'trans' with 'Automatic'/'Manual' names for plotting.

```r
d <- mtcars
d$trans <- factor(d$am, labels=c("Automatic","Manual"))
```

## Exploratory Analysis

Structure of the dataset, with new variable 'trans' included.

```r
str(d)
```

```
## 'data.frame':    32 obs. of  12 variables:
##  $ mpg  : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl  : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp : num  160 160 108 258 360 ...
##  $ hp   : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat : num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt   : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec : num  16.5 17 18.6 19.4 17 ...
##  $ vs   : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am   : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear : num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb : num  4 4 1 1 2 1 4 2 2 4 ...
##  $ trans: Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
```

A boxplot of mpg versus transmission type (see appendix) shows that the median mpg value for cars with manual transmissions is higher than that for cars with automatic transmissions.

## Regression Models

There are 3 models using 3 sets of regressors; transmission type alone, transmission type with engine displacement, and finally transmission type with weight.

The first model is a linear regression of mpg on transmission type (Automatic or Manual). The intercept is the mean mpg for cars with automatic transmissions, the smaller confidence interval indicating more confidence in the mean mpg for automatic compared to manual transmissions. The second estimate (transManual) is the comparison of manual to automatic transmission mpg, the p-value $< 0.05$ shows that the mpg difference is significant at the 95% level.

```r
fit1 <- lm(mpg ~ trans, d)
tidy(fit1)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    17.1       1.12     15.2  1.13e-15
## 2 transManual     7.24      1.76      4.11 2.85e- 4
```

```r
confint(fit1)
```

```
##                2.5 %   97.5 %
## (Intercept) 14.85062 19.44411
## transManual  3.64151 10.84837
```

```r
d %>% filter(trans == 'Automatic') %>% summarise(mean(mpg))
```

```
##   mean(mpg)
## 1  17.14737
```

```
d %>% filter(trans == 'Manual') %>% summarise(mean(mpg))
```

```
##   mean(mpg)
## 1  24.39231
```

The second model considers transmission type and engine displacement which is related to vehicle power performance. When including engine displacement, the difference in mpg between Automatics and Manuals is reduced and not statistically significant (transManual, $p > 0.05$).

```
fit2 <- lm(mpg ~ trans + disp, d)
tidy(fit2)
```

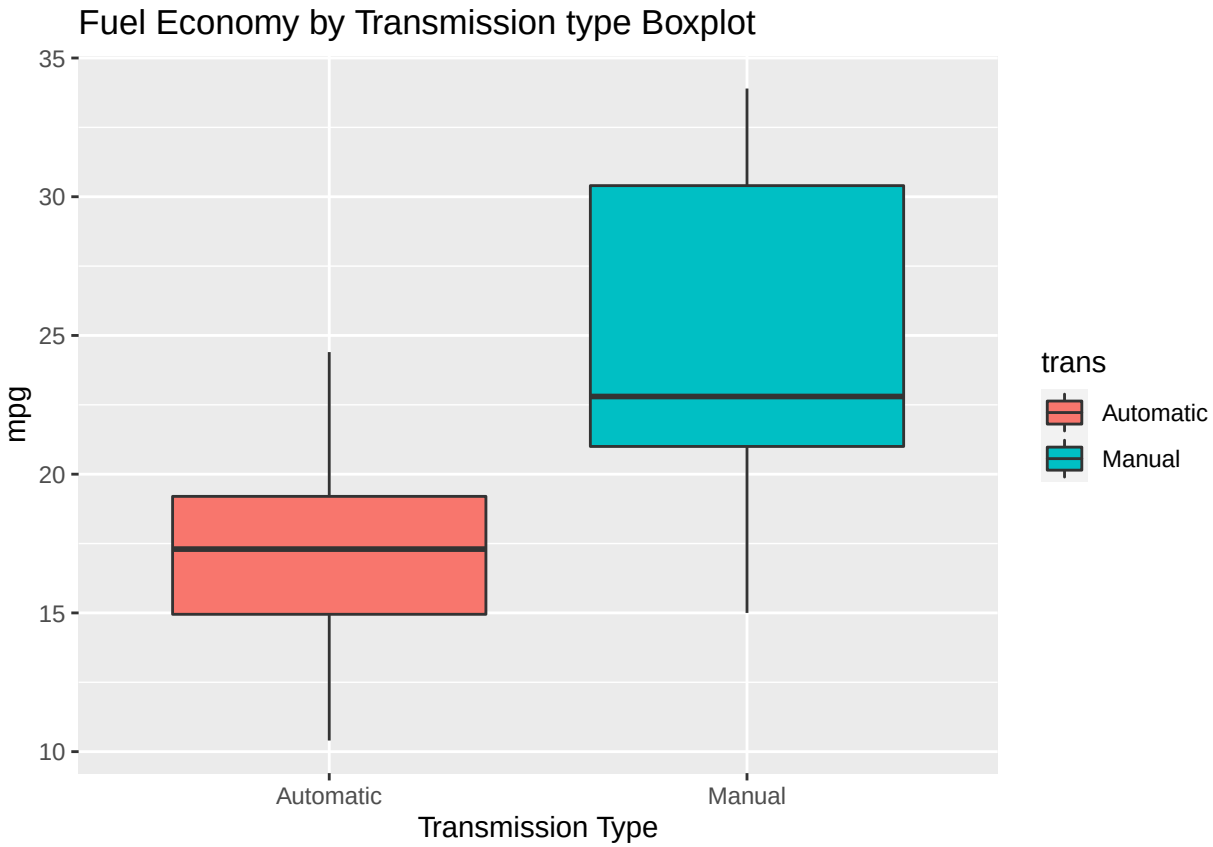
```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)  27.8      1.83        15.2  2.45e-15
## 2 transManual   1.83     1.44         1.28 2.12e- 1
## 3 disp         -0.0369   0.00578     -6.37 5.75e- 7
```

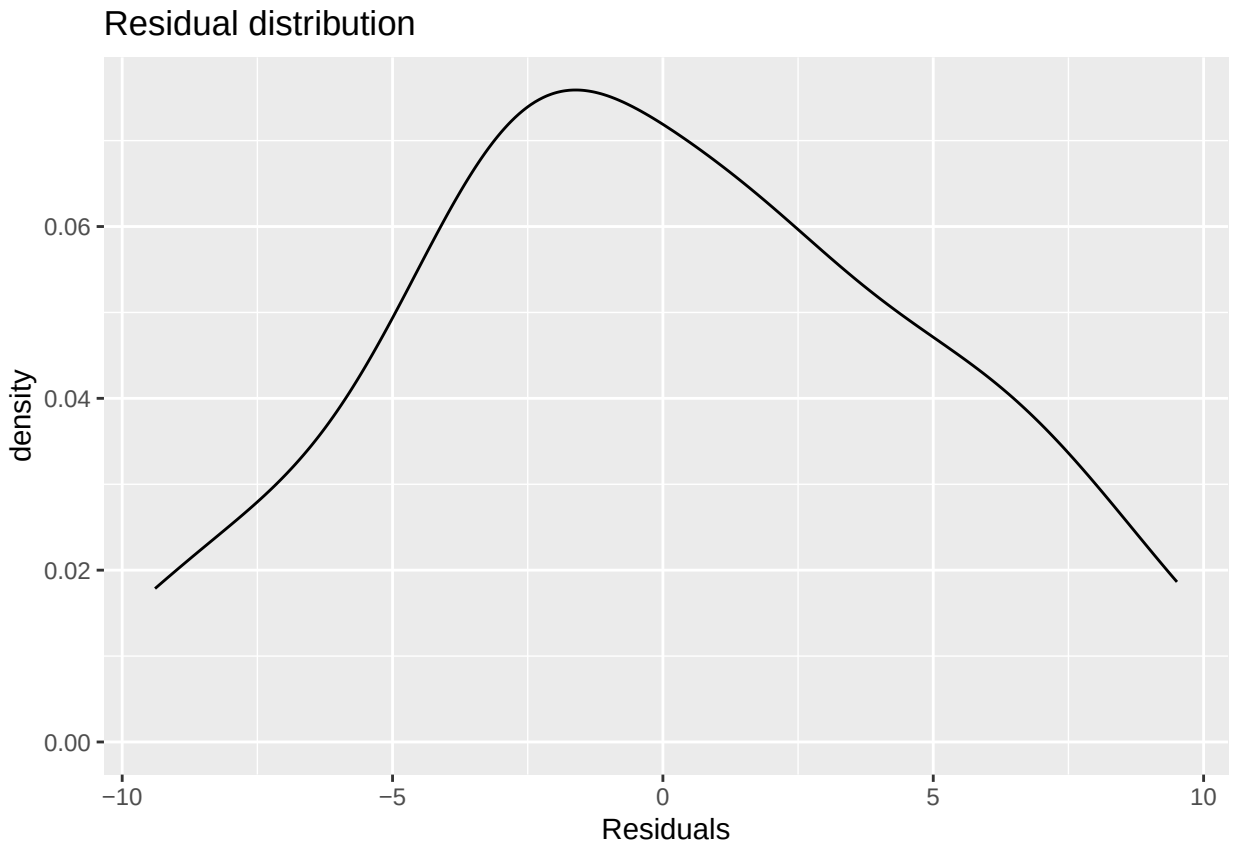
The third model considers transmission type and vehicle weight. When including vehicle weight, the difference in mpg between Automatics and Manuals is reduced and not statistically significant (transManual, $p > 0.05$).

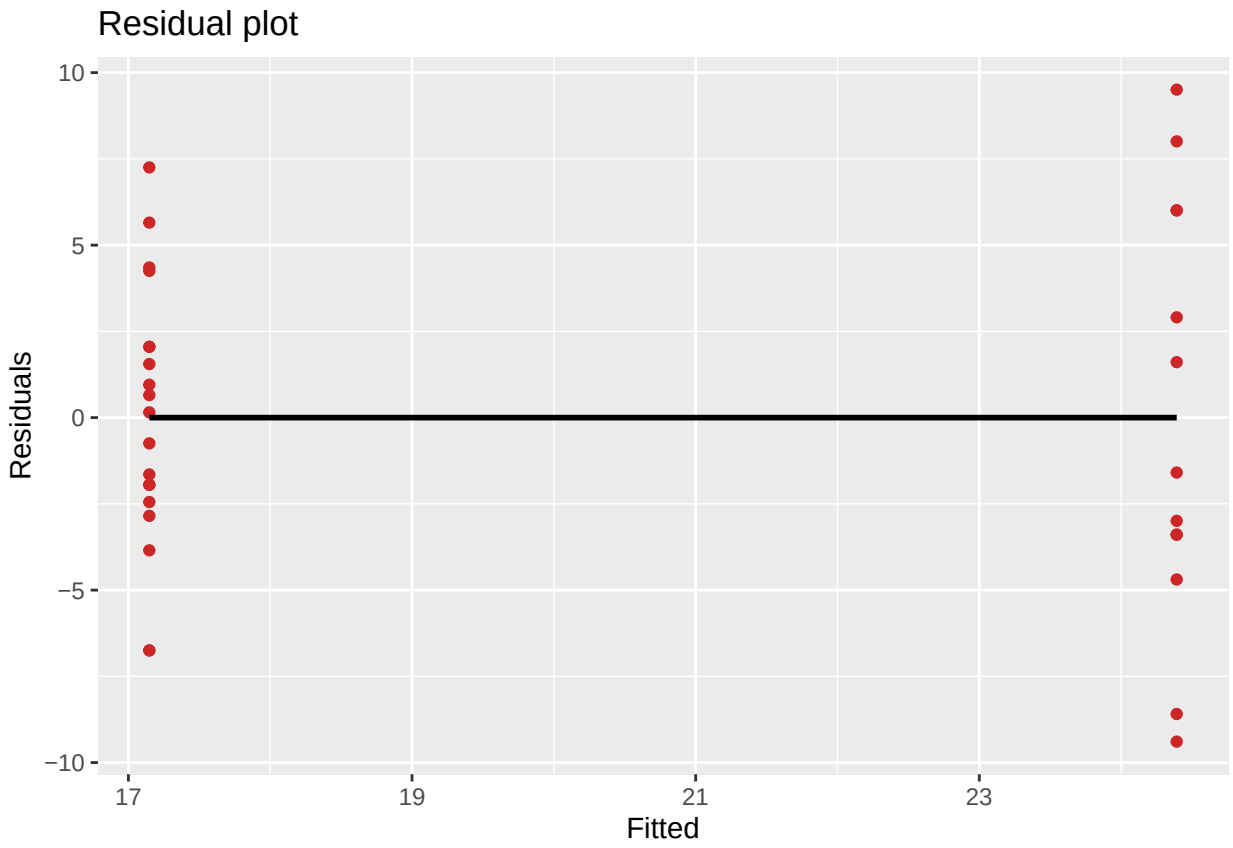
```
fit3 <- lm(mpg ~ trans + wt, d)
tidy(fit3)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)  37.3        3.05    12.2    5.84e-13
## 2 transManual  -0.0236     1.55    -0.0153 9.88e- 1
## 3 wt           -5.35       0.788   -6.79   1.87e- 7
```

## Appendix


Fuel Economy by Transmission type Boxplot
35
30
25
20
15
10
mpg
Automatic
Manual
Transmission Type
trans
Automatic
Manual


Residual distribution
density
0.06
0.04
0.02
0.00
−10
−5
0
5
10
Residuals


Residual plot
Residuals
10
5
0
−5
−10
17
19
21
23
Fitted