The Effect of Transmission on MPG

Steve Myles

21 November 2015

Executive Summary

This report is an investigation of the relationship between transmission (manual vs. automatic) and miles per gallon (MPG) for thirty-two automobiles featured in *Motor Trend* in 1974. Specifically, the following two questions are addressed:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

From the data, it is apparent that manual transmissions are better for MPG (on average, manual transmissions have higher MPG than automatic transmissions). The difference between them, removing confounding variables weight and horsepower, is approx. 2 MPG (average automatic transmission MPG = 34.00 and average manual transmission MPG = 36.09).

Analysis

First, a brief exploratory analysis was performed. This includes the generation of a box and whisker plot to compare the distributions of MPG for each type of transmission (see Appendix, Fig. 1).

From this plot, it is clear that on average and in the majority of cases (due to the small overlap of the ranges) manual transmissions result in higher (better) MPG. Additionally, the spread of MPG for manual transmissions is larger than that for automatic transmissions.

The analysis was continued with a simple linear model, comparing MPG to transmission:

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

"Estimate" in the above table is the mean MPG for each factor, either automatic (17.147, the intercept) or manual (24.39 = 17.147 + 7.245). This confirms what is shown in the plot: manual transmissions do, on average, result in higher MPG (and are therefore better in this regard). The note about the spread of values is also confirmed as MPG for manual transmissions has a higher standard error.
This model's residuals plots can be seen in the Fig. 2 of the Appendix.

However, it should be noted that some of the other variables in the data set may have a confounding effect. This is evidenced by the model's relatively low adjusted R^2 (0.3384589, indicating that only 33.8% of the variation in MPG is explained by the variables in the model). As such, it was necessary to fit other linear models that remove the effects of other variables from consideration in order to determine the true MPG difference between automatic and manual transmissions.

Per the following table which shows the correlations of `mpg` and the remaining nine variables (all variables in the `mtcars` data set with the exception of `am` and `mpg`), it is obvious that several other variables are related to MPG.

```
##            cyl       disp         hp      drat         wt     qsec
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##             vs      gear       carb
## [1,] 0.6640389 0.4802848 -0.5509251
```

For the purposes of this analysis, only those variables with very strong (positive or negative) correlations were considered. Very strong correlations are defined here as +/- 0.7 (definition from here).

This leaves four variables, `cyl`, `disp`, `hp`, and `wt`. The correlation of these variables with each other was then checked to avoid having the final model include variables that are very strongly correlated with each other.

```
##            cyl      disp        hp        wt
## cyl  1.0000000 0.9020329 0.8324475 0.7824958
## disp 0.9020329 1.0000000 0.7909486 0.8879799
## hp   0.8324475 0.7909486 1.0000000 0.6587479
## wt   0.7824958 0.8879799 0.6587479 1.0000000
```

As can be seen in the above table, each of these variables has a very strong correlation with the others (using the +/- 0.7 definition as described above), with the exception of horsepower (`hp`) and weight in thousands of pounds (`wt`). As such, only these two variables were considered for inclusion into the linear model.

First, `wt` was added to the linear model resulting in an adjusted R^2 of 0.7357889. This is much improved over the original adjusted R^2 of 0.3384589. `hp` was then added to the model, resulting in an adjusted R^2 of 0.8227357 (indicating that 82.3% of the variation in MPG is explained by the variables in the model). As this has the highest adjusted R^2 of the models considered, this is the best representation of the effect of transmission type on MPG.

The model's coefficients are:

```
##                   Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept)    34.00287512 2.642659337 12.866916 2.824030e-13
## as.factor(am)1  2.08371013 1.376420152  1.513862 1.412682e-01
## wt             -2.87857541 0.904970538 -3.180850 3.574031e-03
## hp             -0.03747873 0.009605422 -3.901830 5.464023e-04
```

"Estimate" in the above table is the effect of each variable on mean MPG, either automatic transmission (34.00, the intercept) or manual transmission (36.09 = 34.003 + 2.084). This confirms the earlier assessment that manual transmissions on average result in higher MPG, though the difference between them is much less pronounced when the confounding effects of weight and horsepower are removed.
Weight and horsepower both have negative effects on mean MPG, which makes logical sense - more powerful (higher horsepower) and/or heavier (higher weight) engines would require more fuel to operate and thus would have lower gas mileage (MPG). This model's residuals plots can be seen in the Fig. 3 of the Appendix.

Note

This report was written in R Markdown and converted to pdf via knitr. For the source .Rmd document, please click here.

Appendix

Fig. 1

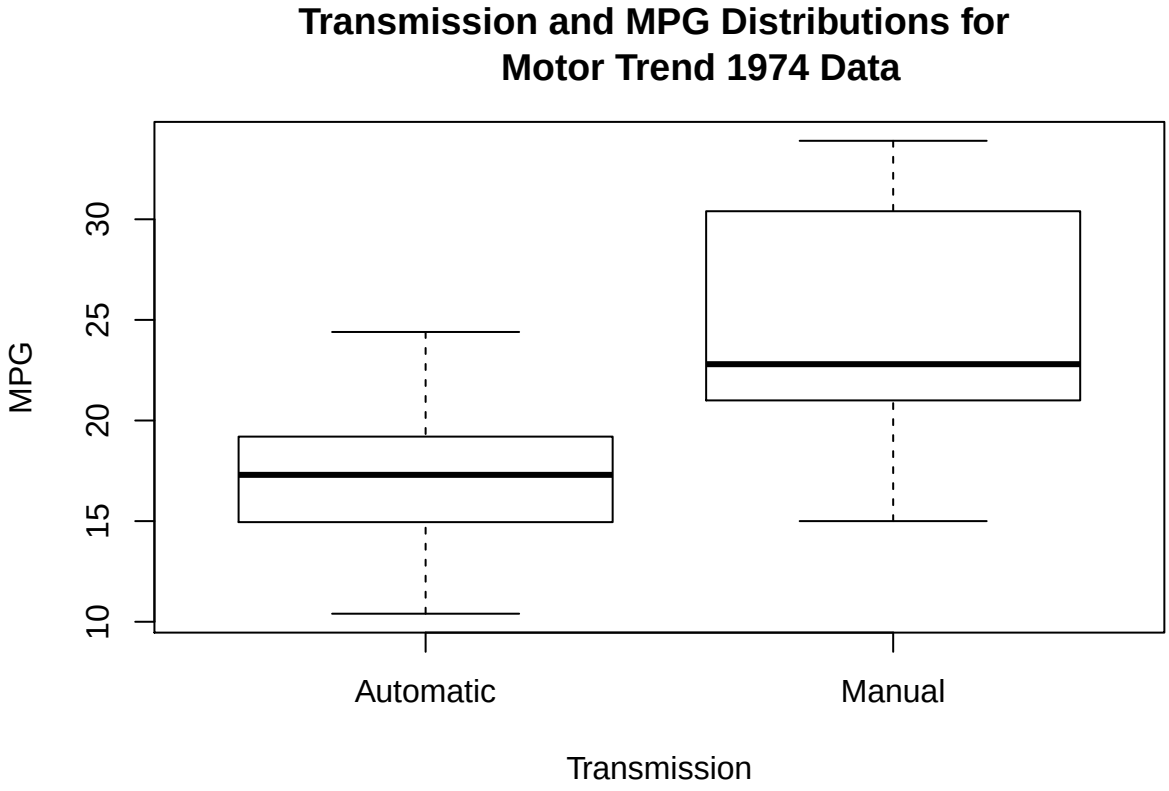


Fig. 2

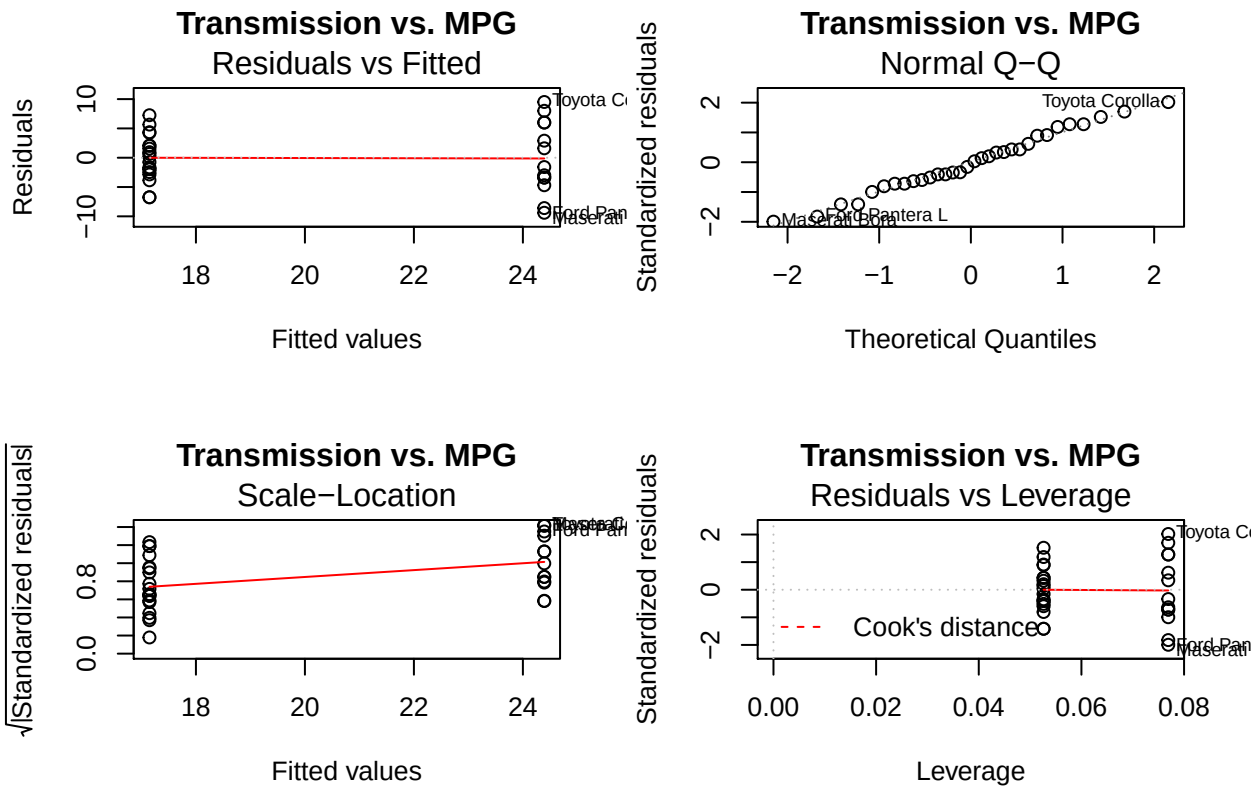
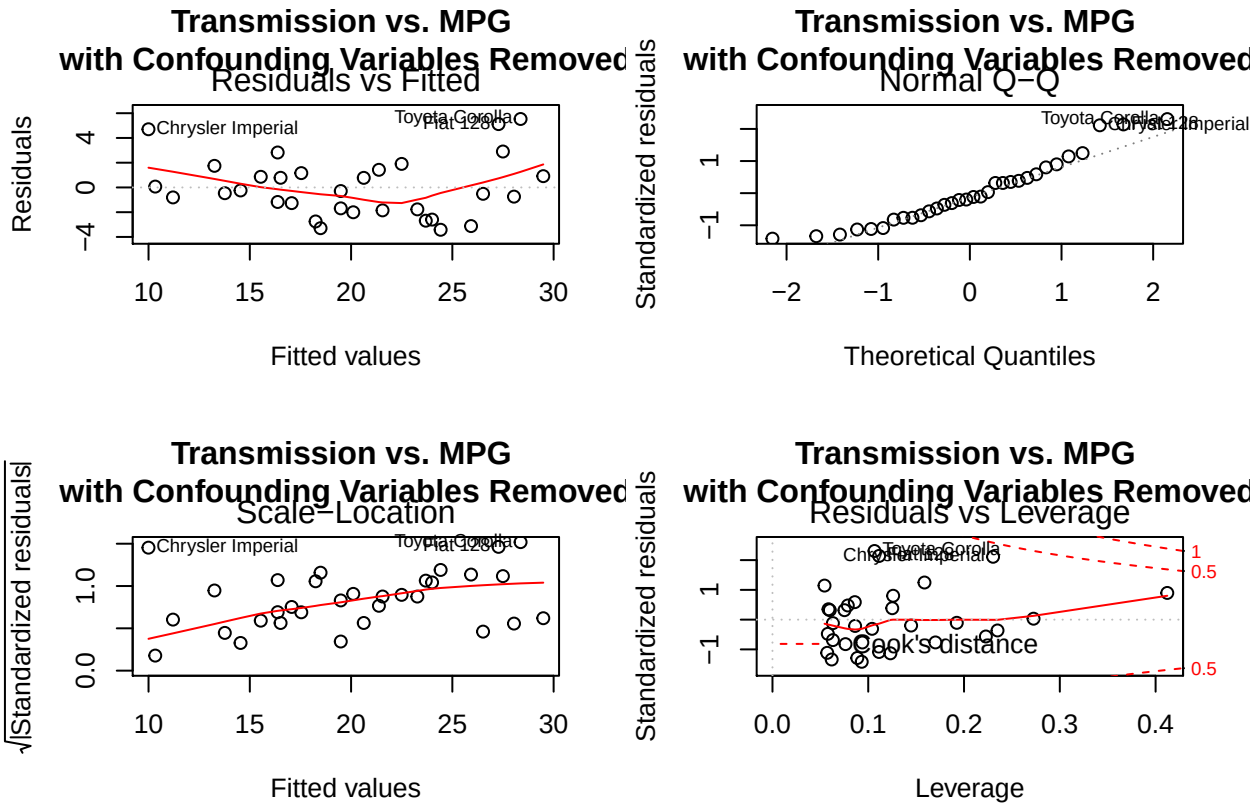

Transmission vs. MPG
Residuals vs Fitted
Residuals
Fitted values
Transmission vs. MPG
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Transmission vs. MPG
Scale−Location
√|Standardized residuals|
Fitted values
Transmission vs. MPG
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

Fig. 3


Transmission vs. MPG
with Confounding Variables Removed
Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Transmission vs. MPG
with Confounding Variables Removed
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Transmission vs. MPG
with Confounding Variables Removed
Scale−Location
√|Standardized residuals|
Fitted values
Transmission vs. MPG
with Confounding Variables Removed
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage