Transmission Type Vs. MPG

Austin L. Bistline

May 30, 2018

Summary

This document seeks to answer the question of which transmission type is better for miles per gallon(mpg). An exploration of the 'mtcars' dataset shows a comparison of 32 different cars with 11 different characteristics or variables. Several models are then created to determine which variables show a significant ability to determine fuel mileage. The final model indicates that when weight and quarter mile time are considered, cars with manual transmissions get better mileage, but only by 3.7%. This estimate contrasts a comparison of the two car types by average mileage which shows that cars with manual transmissions get better mileage by 42.7%. The difference can be explained by a comparison of the average weights between car types, which shows that cars with automatic transmissions are also 58% heavier.

Introduction

In 1974 data originated from Motor Trend, a magazine about the automobile industry. From that dataset we explore the relationship between a set of variables and miles per gallon (MPG) (outcome) with a particular interest in answering the following two questions:

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

This report is the result of a peer-graded assignment, the Regression Models course project created by Johns Hopkins University and distributed through Coursera.

Exploring the Data

We first will become familiar with the data by checking the dimensions of the dataset and taking a look at the variables, then first 3 rows.

```
## [1] 32 11
```

```
##                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
```

There are 32 different cars and 11 different characteristics. We will only consider mileage (mpg), engine size (disp), horsepower (hp), weight (wt), and quarter mile time (qsec) by transmission type (am).

```
# Pulling variables out of mtcars for simplicity, and also creating a factor variable "trans"
trans = as.factor(ifelse(mtcars$am==0, "Automatic", "Manual"))
mpg = mtcars$mpg; disp = mtcars$disp; hp = mtcars$hp; wt = mtcars$wt; qsec = mtcars$qsec
```

Model Fitting to Predict MPG

A model with all variables will be created to determine which ones are statistically significant. Variables will a P-value greater than 0.05 will be considered insignificant for predicting mpg. Statistically insignificant variables will be removed until a model has only significant variables. The P-values are extracted in the following code:

```
fit1 = lm(mpg ~ disp + hp + wt + qsec + trans -1)
# P-values of fit1
c(summary(fit1)$coef[, 4], sigma = sigma(fit1), R.squared=summary(fit1)$r.squared)
```

```
##           disp             hp             wt           qsec transAutomatic
##    0.298972150    0.156387279    0.002075008    0.043907652    0.152378367
##    transManual          sigma      R.squared
##    0.057628233    2.429290863    0.989073179
```

In this model, the disp variable is the least statistically significant, and hp is statistically insignificant as well. The next model will remove only disp and the P-values are again shown in the following:

```
fit2 = lm(mpg ~ hp + wt + qsec + trans -1)
# P-values of fit2
c(summary(fit2)$coef[, 4], sigma=sigma(fit2), R.squared=summary(fit2)$r.squared)
```

```
##             hp             wt           qsec transAutomatic    transManual
##    0.223087932    0.001141407    0.075731202    0.072149342    0.026442873
##          sigma      R.squared
##    2.434828005    0.988601130
```

We see that the removal of disp did not affect hp, and hp is still statistically insignificant. Therefore hp will be removed as well. P-values shown below:

```
fit3 = lm(mpg ~ wt + qsec + trans -1)
# P-values of fit3
c(summary(fit3)$coef[, 4], Sigma=sigma(fit3), R.squared=summary(fit3)$r.squared)
```

```
##             wt           qsec transAutomatic    transManual          Sigma
##   6.952711e-06   2.161737e-04   1.779152e-01   4.754335e-02   2.458846e+00
##      R.squared
##   9.879446e-01
```

This model shows that when grouped separately, both transmission types are less statistically significant. A wt^2 term will be added to see if the model can be improved. The following scripts show the P-values first, then the coefficient estimates:

```
fit4 = lm(mpg ~ wt + I(wt^2) + qsec + trans -1)
# P-values of fit4
summary(fit4)$coef[, 4]
```

```
##             wt        I(wt^2)           qsec transAutomatic    transManual
##   0.0002628587   0.0089778354   0.0014614032   0.0045281885   0.0011471599
```

```
# Coefficient Estimates of fit4
c(summary(fit4)$coef[, 1], Sigma=sigma(fit4), R.squared=summary(fit4)$r.squared)
```

```
##             wt        I(wt^2)           qsec transAutomatic    transManual
##    -11.2490694      0.9581950      0.9705112     27.7376147     28.7591331
##          Sigma      R.squared
##      2.2014922      0.9906812
```

With the wt^2 term in the model, all variable terms are statistically significant. This model shows that mpg for automatics and manuals are very similar with a difference of only 1 mpg, or 3.8%, at the intercept.

```
##   Transmission  MPG  Disp   WT    HP Qsec
## 1    Automatic 17.1 290.4 3800 160.3 18.2
## 2       Manual 24.4 143.5 2400 126.8 17.4
```

Our final model shows that despite the average mpg being nearly 43% better in manuals, actual mpg is very similar when we take weight and performance into consideration. The residual standard error for the model "fit4" is quite high at 2.201, and the 95% confidence interval shows that mpg could vary from 33% less efficient to 2.4% more efficient in automatics.The small number of cars used in this analysis increases the model error with a broad-ranging confidence interval shown below:

```
confint(fit4, level=.95)
```

```
##                       2.5 %    97.5 %
## wt               -16.7495203 -5.748619
## I(wt^2)            0.2599863  1.656404
## qsec               0.4085024  1.532520
## transAutomatic     9.3584517 46.116778
## transManual       12.5335925 44.984674
```

Appendix

```
par(mfrow = c(1, 2))
plot(fit3, which = 1)
plot(fit4, which = 1)
```

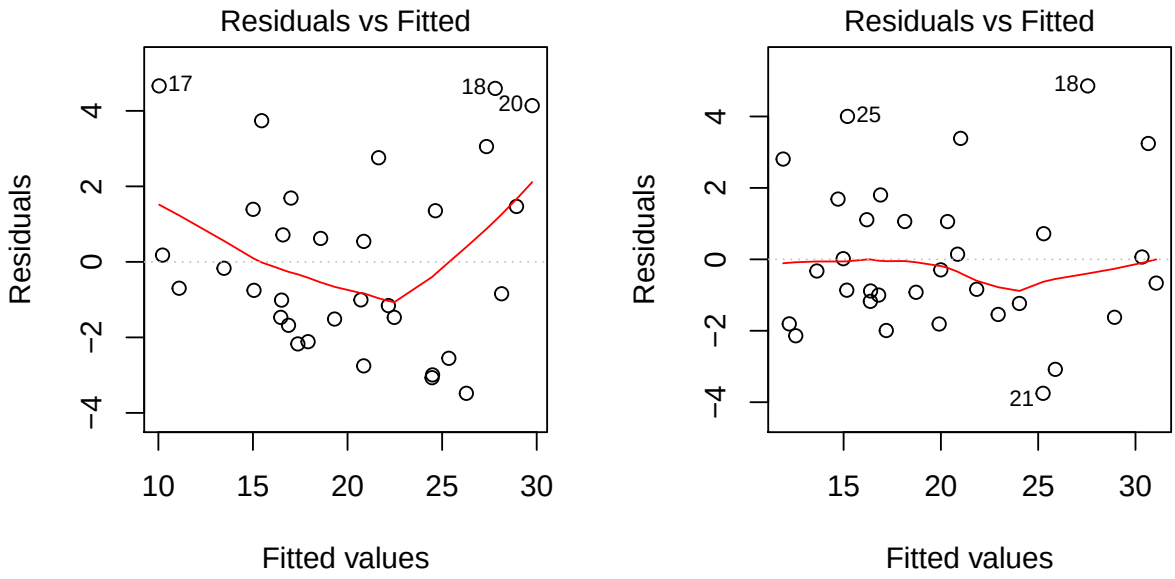

Figure1: Residual plot of model 'fit3' (left) and model 'fit4' (right). In 'fit3' we see that the residuals are non-linear, which prompted the inclusion of a wt^2 term.