Is Transmission type better for MPG

Aaron Augustine

August 11, 2015

Executive Summary

The purpose of this paper is to explore the relationship between a set of variables and miles per gallon(mpg) as part of a class project using the mtcars dataset. It will focus on answering two questions: "Is an automatic or manual transmission better for MPG?" and "Quantify the MPG difference between automatic and manual transmissions". Based on the analysis below there is a relationsip between MPG and transmission type.

Data Analysis

EDA

First we'll start by exploring relationships in the data. We'll do this by creating a correlation matrix and pairwise plots of all of the variabes. (See appendix). Interestingly enough all the variables seem to have strong correlations with mpg. It's also noted that we will need to convert several of the variables from numeric to factor variables prior to feeding into a regression model.

From here we'll work to build a regression model to determine determine how Transmission (0 = automatic, 1 = manual) relates to MPG.

Variable Selection and Model Building

To get started with variable selection we execute a simple model of transmission, am, vs. mpg. (See the Appendix for the model summary)

```
fit1 <- lm(mpg~am,data=regdata)
```

The model is significant with a p-value of 0.000285020744. However, given the lower R-squared 0.33846 we'll work to include more predictors to get a better explaination of the variation in mpg. To do this we'll run a stepwise regression and compare the models. (See the Appendix for the model summary)

```
fit2 <- lm(mpg~cyl+disp+hp+drat+wt+qsec+vs+am+gear+carb,data=regdata)
stepModel <- step(fit2, k=log(nrow(regdata)))
```

From the stepwise regression, we find a signifcant model with a pvalue of 1.2e-11 and a much better R-squared, 0.83356. Comparing the two models using the anova function, we can see that it is benefical to add in wt and qsec as we obtain a signifcant pvalue from the anova. (see below)

```
anova(fit1,stepModel)
```

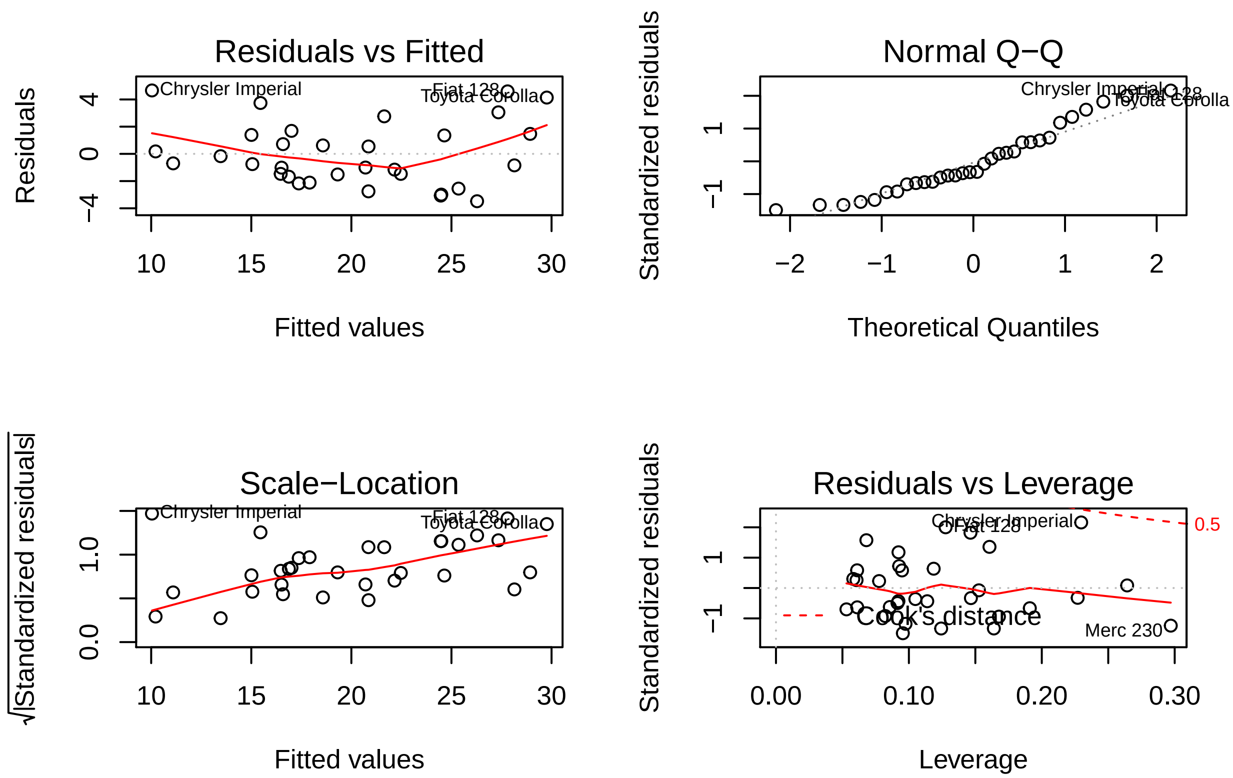
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
```

```
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     28 169  2       552 45.6 1.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnostic Plots

With a model selected lets do some quick diagnostic plots.

```r
par(mfrow=c(2,2))
plot(stepModel)
```



Looking at the plot of the fit values vs. the residuals, there does not seem to be any strong heteroskedasticity and the errors center around zero. One might argue there is some curve to the errors so additional variables might be helpful but its difficult to tell with a smaller sample. The Normal Q-Q plot also doesn't show any strong depatures from normality. The Scale Location and Residuals vs. Leverage don't show any outliers.

Conclusion

Lastly, to answer the given questions, we can execute a T-test on the transmission parameter,am, from the stepwise model. From this we can find that the paramater is moderately signifcant with a Pvalue of, 0.04672. Thus, a manual transmission provides 2.93584 more miles per gallon keeping in mind there is a lot variation around the estimate. The 95% CI for the interval is [0.04573,5.82594].

Appendix

Model Summary

```r
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = regdata)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## am1             7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

```r
summary(stepModel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = regdata)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.481 -1.556 -0.726  1.411  4.661
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.618      6.960    1.38  0.17792
## wt            -3.917      0.711   -5.51    7e-06 ***
## qsec           1.226      0.289    4.25  0.00022 ***
## am1            2.936      1.411    2.08  0.04672 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.46 on 28 degrees of freedom
## Multiple R-squared:  0.85,   Adjusted R-squared:  0.834
## F-statistic: 52.7 on 3 and 28 DF,  p-value: 1.21e-11
```

EDA & Plots

```r
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
```

```r
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 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 ...
```

```r
cor(mtcars)
```

```
##           mpg      cyl     disp       hp      drat       wt      qsec
## mpg   1.00000 -0.85216 -0.84755 -0.77617  0.681172 -0.86766  0.418684
## cyl  -0.85216  1.00000  0.90203  0.83245 -0.699938  0.78250 -0.591242
## disp -0.84755  0.90203  1.00000  0.79095 -0.710214  0.88798 -0.433698
## hp   -0.77617  0.83245  0.79095  1.00000 -0.448759  0.65875 -0.708223
## drat  0.68117 -0.69994 -0.71021 -0.44876  1.000000 -0.71244  0.091205
## wt   -0.86766  0.78250  0.88798  0.65875 -0.712441  1.00000 -0.174716
## qsec  0.41868 -0.59124 -0.43370 -0.70822  0.091205 -0.17472  1.000000
## vs    0.66404 -0.81081 -0.71042 -0.72310  0.440278 -0.55492  0.744535
## am    0.59983 -0.52261 -0.59123 -0.24320  0.712711 -0.69250 -0.229861
## gear  0.48028 -0.49269 -0.55557 -0.12570  0.699610 -0.58329 -0.212682
## carb -0.55093  0.52699  0.39498  0.74981 -0.090790  0.42761 -0.656249
##            vs        am     gear      carb
## mpg   0.66404  0.599832  0.48028 -0.550925
## cyl  -0.81081 -0.522607 -0.49269  0.526988
## disp -0.71042 -0.591227 -0.55557  0.394977
## hp   -0.72310 -0.243204 -0.12570  0.749812
## drat  0.44028  0.712711  0.69961 -0.090790
## wt   -0.55492 -0.692495 -0.58329  0.427606
## qsec  0.74454 -0.229861 -0.21268 -0.656249
## vs    1.00000  0.168345  0.20602 -0.569607
## am    0.16835  1.000000  0.79406  0.057534
## gear  0.20602  0.794059  1.00000  0.274073
## carb -0.56961  0.057534  0.27407  1.000000
```

```r
pairs(mtcars, panel=panel.smooth)
```