Regression Week 4 Project

Michael Kroog

January 13, 2018

Executive Summary

Motor Trend magazine is interested in exploring the relationship between a set of variables and miles per gallon. We are specifically looking to answer the following two questions:

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

After some exploratory data analysis to get an idea of the data we are working with and looking at some plots we will see that a manual transmission is better for MPG. The mean MPG for an automatic transmission is 17.15 and the mean MPG for a manual is 24.39 with a difference of 7.24.

However since other variables effect the MPG of a car we should examine the other 9 variables in the data set. After conducting a backward stepwise algorithm we see that the variables wt and qsec may also be significant in predicting the MPG of a car. To test this I ran an ANOVA to determine if the addition of wt and qsec were significant. I set the Null Hypothesis to "the models do not significantly differ" and the Alternative Hypothesis to "the model with the addition of wt and qsec is better". The result of the p-value from the ANOVA is 1.55e-09 at the 0.001 level, which means we do not fail to reject the null and conclude that we have evidence to believe the model with wt and qsec is better.

Now that we have a model that appears to be statistically better lets take a look at some diagnostic plots. The Residuals vs Fitted plot looks fairly random and doesn't show any systematic model departure. The Q-Q plot looks normal. The Scale-Location plot however does show some evidence of heteroskedasticity. The Residuals vs Leverage plot show a small Cook's distance lines with no points outside of the lines, this doesn't show any cases with large influence.

```r
head(mtcars, 4); str(mtcars); summary(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
```

```
## '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 ...
```

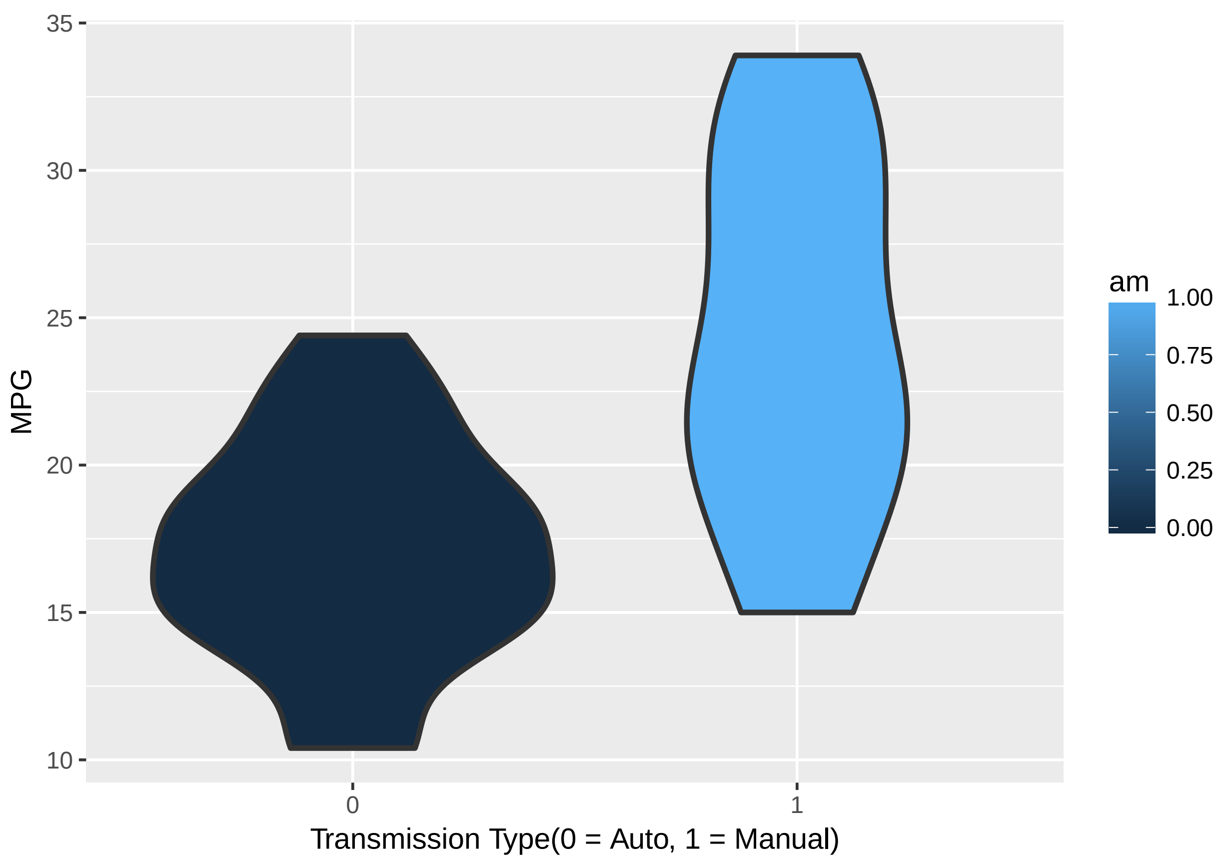
```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
```

```
##  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
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```r
g <- ggplot(mtcars, aes(factor(am), mpg, fill = am))
g <- g + geom_violin(size = 1)
g <- g + xlab("Transmission Type(0 = Auto, 1 = Manual)") + ylab("MPG")
g
```



```r
mean(mtcars$mpg[mtcars$am == 0]) #mean of automatic
```

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

```r
mean(mtcars$mpg[mtcars$am == 1]) #mean of manual
```

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

```r
mean(mtcars$mpg[mtcars$am == 1]) - mean(mtcars$mpg[mtcars$am == 0]) #difference of means
```

```
## [1] 7.244939
```

```r
summary(lm(mpg ~ factor(am), mtcars))$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

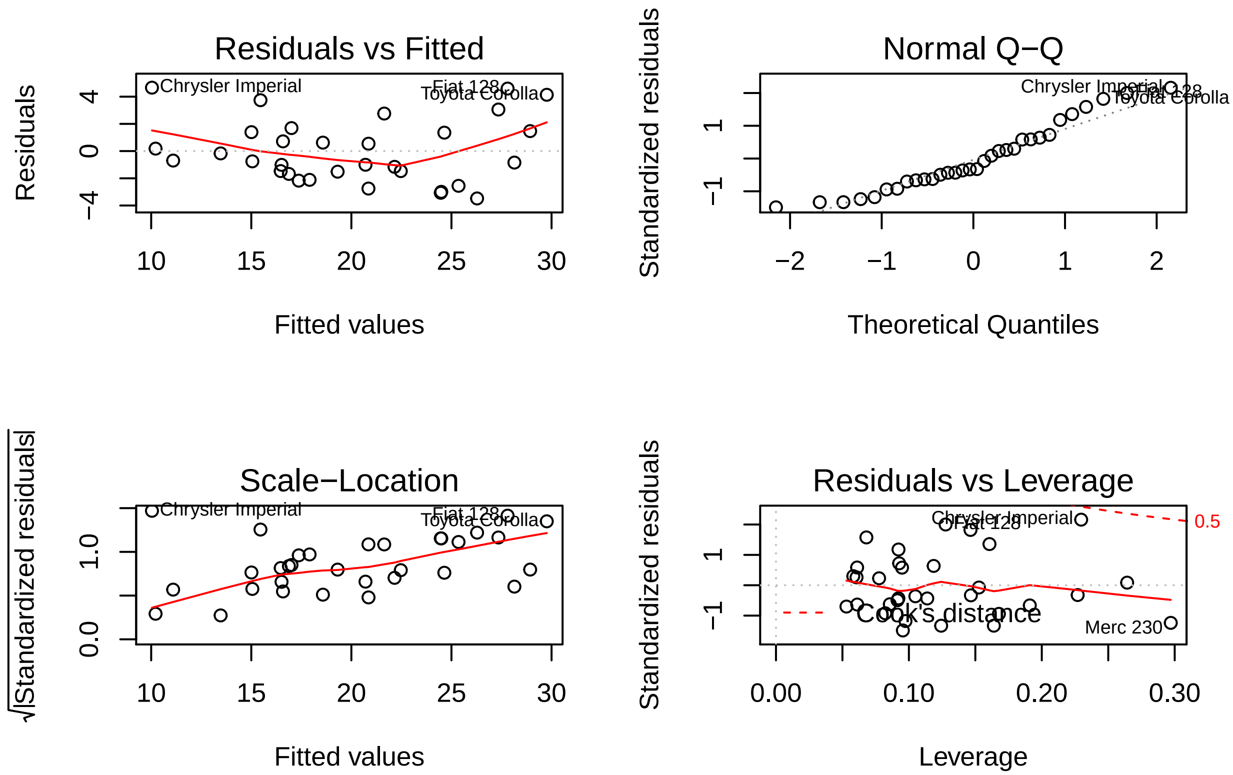
```r
fit1 <- lm(mpg ~ am, mtcars)

# step(lm(mpg ~ ., mtcars)) stepwise algorithm, printing takes up too much space

fit2 <- update(fit1, mpg ~ am + wt + qsec)
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow = c(2, 2))
plot(fit2)
```



Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
Merc 230
0.5

‘‘‘