MPG for Automatic and Manual Transmission

Ramana Sonti

3/4/2017

Executive Summary (Overview): This report analyzes Motor Trend data to figure out if the manual or automatic transmission gets better MPG. In addition, it tries to quantifies the MPG difference between automatic and manual transmission.

Exploratory Data Analysis:

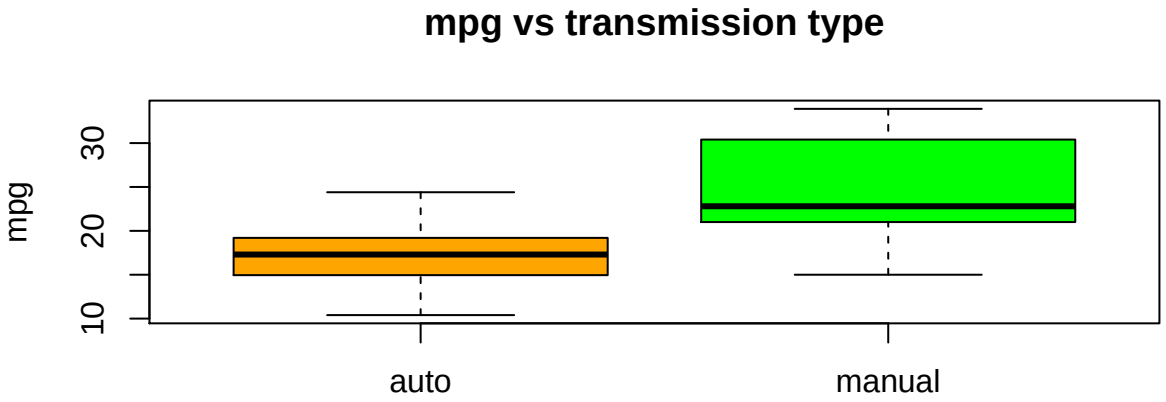
```
library(datasets)
library(stats)
mtcars2 <- mtcars
head(mtcars2, 2)
```

```
##               mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4      21   6  160 110  3.9 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21   6  160 110  3.9 2.875 17.02  0  1    4    4
```

```
str(mtcars2)
```

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

```
boxplot(mpg ~ am, data = mtcars2, ylab="mpg", main="mpg vs transmission type", names=c("auto", "manual"
```



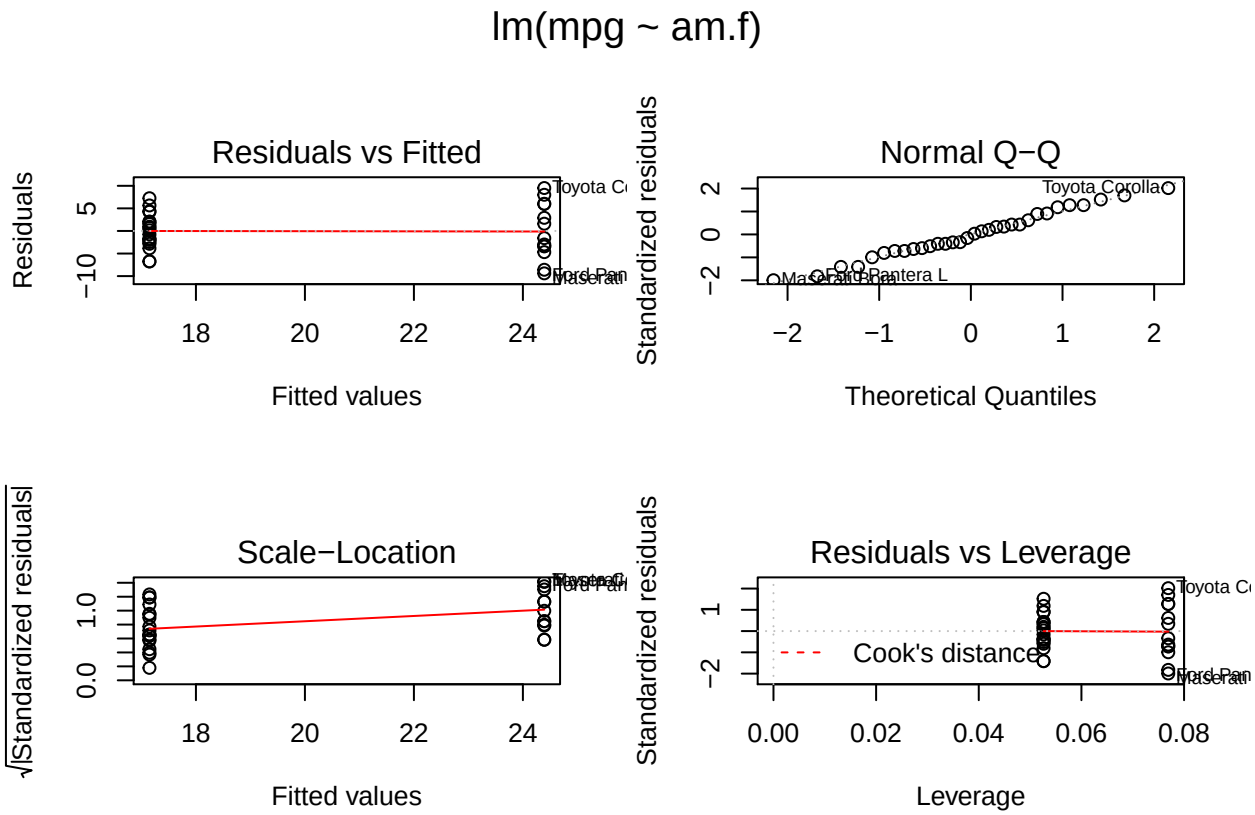
Modeling Options:

The box plot for the data above shows that MPG for manual transmission is better. We will model MPG as outcome and am as predictor and see if it makes sense in predicting the outcomes when am is considered alone as a predictor. We will consider models with all other variables with and without am for the sake of comparison.

Please note that glm will not be considered as a model fitting option here as the outcome is neither binary nor poisson.

Residual Plots and Diagnostics:

```
mtcars2$am.f <- as.factor(mtcars2$am)
levels(mtcars2$am.f) <- c("Auto", "Manual")
lmfit <- lm(mpg ~ am.f, data = mtcars2)
par(mfrow = c(2, 2), oma = c(0, 0, 2, 0))
plot(lmfit)
```

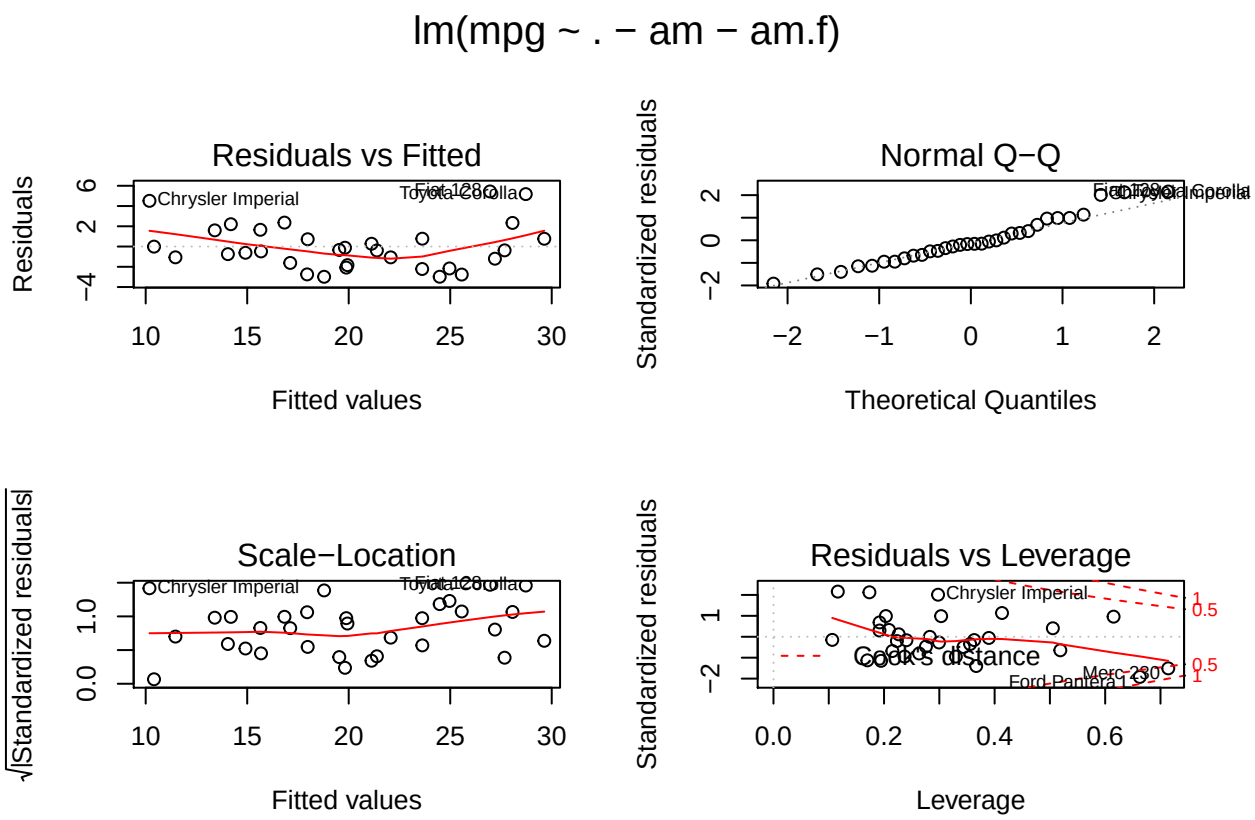



```
summary(lmfit)
```

```
##
## Call:
## lm(formula = mpg ~ am.f, data = mtcars2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am.fManual     7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

```
lmfit0 <- lm(mpg ~ . -am -am.f, data = mtcars2)
lmfit1 <- lm(mpg ~ . -am, data = mtcars2)
plot(lmfit0)
```




```
summary(lmfit0)
```

```
##
## Call:
## lm(formula = mpg ~ . - am - am.f, data = mtcars2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9886 -1.6738 -0.3834  0.9796  5.4395
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.59704   18.41902   0.955   0.3498
## cyl         -0.44543    1.02029  -0.437   0.6667
```

```
## disp         0.01275    0.01805   0.706   0.4876
## hp          -0.02022    0.02199  -0.919   0.3679
## drat         1.10054    1.63356   0.674   0.5075
## wt          -3.93430    1.90734  -2.063   0.0511 .
## qsec         0.57571    0.71086   0.810   0.4267
## vs          -0.22311    2.08103  -0.107   0.9156
## gear         1.22953    1.43393   0.857   0.4004
## carb        -0.26242    0.83653  -0.314   0.7567
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.68 on 22 degrees of freedom
## Multiple R-squared:  0.8596, Adjusted R-squared:  0.8022
## F-statistic: 14.97 on 9 and 22 DF,  p-value: 1.855e-07
```

Model Selection:

```
anova(lmfit, lmfit0)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am.f
## Model 2: mpg ~ (cyl + disp + hp + drat + wt + qsec + vs + am + gear +
##     carb + am.f) - am - am.f
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1     30 720.90
## 2     22 158.04  8    562.86 9.794 1.045e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
comparedf <- cbind(mtcars2[c("mpg","am.f","wt")]
            ,data.frame(round(predict(lmfit),1))
            ,data.frame(round(predict(lmfit0),1))
            ,data.frame(round(predict(lmfit1),1))
            ,data.frame(round(resid(lmfit),2))
            ,data.frame(round(resid(lmfit0),2))
            ,data.frame(round(resid(lmfit1),2))
                    )
names(comparedf) <- c("mpg","am.f","wt","fit","fit0","fit1","resid","resid0","resid1")
```

Comparison of predicted values from models:

- fit = predicted mpg for the model with am as the only predictor
- fit0 = predicted mpg for the model with all other variables as predictors without am (am.f)
- fit1 = predicted mpg for the model with all other variables including am (am.f) as predictors
- resid, resid0, and resid1 are corresponding residuals.
- Also listed wt to show that other variables do have impact on MPG for vehicles with both automatic and manual transmission.

```
comparedf
```

```
##                     mpg   am.f    wt  fit fit0 fit1 resid resid0 resid1
## Mazda RX4          21.0 Manual 2.620 24.4 22.1 22.6 -3.39  -1.07  -1.60
```

```
## Mazda RX4 Wag       21.0 Manual 2.875 24.4 21.4 22.1 -3.39  -0.39  -1.11
## Datsun 710          22.8 Manual 2.320 24.4 25.6 26.3 -1.59  -2.77  -3.45
## Hornet 4 Drive      21.4   Auto 3.215 17.1 21.1 21.2  4.25   0.28   0.16
## Hornet Sportabout   18.7   Auto 3.440 17.1 18.0 17.7  1.55   0.72   1.01
## Valiant             18.1   Auto 3.460 17.1 19.9 20.4  0.95  -1.84  -2.28
## Duster 360          14.3   Auto 3.570 17.1 14.9 14.4 -2.85  -0.61  -0.09
## Merc 240D           24.4   Auto 3.190 17.1 23.6 22.5  7.25   0.77   1.90
## Merc 230            22.8   Auto 3.150 17.1 25.0 24.4  5.65  -2.16  -1.62
## Merc 280            19.2   Auto 3.440 17.1 19.5 18.7  2.05  -0.33   0.50
## Merc 280C           17.8   Auto 3.440 17.1 19.9 19.2  0.65  -2.08  -1.39
## Merc 450SE          16.4   Auto 4.070 17.1 14.2 14.2 -0.75   2.21   2.23
## Merc 450SL          17.3   Auto 3.730 17.1 15.6 15.6  0.15   1.65   1.70
## Merc 450SLC         15.2   Auto 3.780 17.1 15.7 15.7 -1.95  -0.48  -0.54
## Cadillac Fleetwood  10.4   Auto 5.250 17.1 11.5 12.0 -6.75  -1.06  -1.63
## Lincoln Continental 10.4   Auto 5.424 17.1 10.4 10.9 -6.75  -0.01  -0.54
## Chrysler Imperial   14.7   Auto 5.345 17.1 10.2 10.5 -2.45   4.51   4.21
## Fiat 128            32.4 Manual 2.200 24.4 27.0 27.8  8.01   5.44   4.63
## Honda Civic         30.4 Manual 1.615 24.4 29.6 29.9  6.01   0.77   0.50
## Toyota Corolla      33.9 Manual 1.835 24.4 28.7 29.5  9.51   5.18   4.39
## Toyota Corona       21.5   Auto 2.465 17.1 24.5 23.6  4.35  -2.98  -2.14
## Dodge Challenger    15.5   Auto 3.520 17.1 17.1 16.9 -1.65  -1.62  -1.44
## AMC Javelin         15.2   Auto 3.435 17.1 18.0 17.7 -1.95  -2.75  -2.53
## Camaro Z28          13.3   Auto 3.840 17.1 14.0 13.3 -3.85  -0.75  -0.01
## Pontiac Firebird    19.2   Auto 3.845 17.1 16.8 16.7  2.05   2.36   2.51
## Fiat X1-9           27.3 Manual 1.935 24.4 27.7 28.3  2.91  -0.38  -0.99
## Porsche 914-2       26.0 Manual 2.140 24.4 27.2 26.2  1.61  -1.20  -0.15
## Lotus Europa        30.4 Manual 1.513 24.4 28.1 27.6  6.01   2.33   2.76
## Ford Pantera L      15.8 Manual 3.170 24.4 18.8 18.9 -8.59  -2.99  -3.07
## Ferrari Dino        19.7 Manual 2.770 24.4 19.8 19.7 -4.69  -0.12   0.01
## Maserati Bora       15.0 Manual 3.570 24.4 13.4 13.9 -9.39   1.60   1.06
## Volvo 142E          21.4 Manual 2.780 24.4 23.6 24.4 -2.99  -2.22  -2.97
```

Confidence Intervals for model with as as the only predictor:

```
confint(lmfit)
```

```
##                2.5 %   97.5 %
## (Intercept) 14.85062 19.44411
## am.fManual   3.64151 10.84837
```

Confidence Intervals for model with all variables including am:

```
confint(lmfit1)
```

```
##                    2.5 %     97.5 %
## (Intercept) -26.62259745 51.22934576
## cyl          -2.28468553  2.06180457
## disp         -0.02380146  0.05047194
## hp           -0.06675236  0.02378812
## drat         -2.61383350  4.18805545
## wt           -7.65495413  0.22434628
## qsec         -0.69883421  2.34091571
```

```
## vs           -4.05880242  4.69432805
## gear         -2.44999107  3.76081711
## carb         -1.92290442  1.52406591
## am.fManual   -1.75681208  6.79726585
```

Interpreting Coefficients:

The 0.05 confidence interval for the mdoel with am as the only predictor suggests that there is 3.6 to 10.8mpg gain for vehicles with manual transmission. The variance could be mostly attributed to the influence of other variables like wt.

Coefficeints for model with am as only predictor:

```
coef(lmfit)
```

```
## (Intercept)  am.fManual
##   17.147368    7.244939
```

Coefficeints for model with all other variables as predictor including am:

```
coef(lmfit1)
```

```
## (Intercept)         cyl        disp          hp        drat          wt
## 12.30337416 -0.11144048  0.01333524 -0.02148212  0.78711097 -3.71530393
##        qsec          vs        gear        carb  am.fManual
##  0.82104075  0.31776281  0.65541302 -0.19941925  2.52022689
```

Conclusion: When considered alone as a predictor, the model (fit) predicts that the vehicles with manual transmission (24.4) perform better than with automatic (17.1). They don't compare very well with many of the observed values. However, the predicted values for the model (fit0, fit1) with all other variables considered as predictors with or without am are very close to the observed values. The contribution from manual transmission is about 2.5mpg when all variables are considered and is about 7.2mpg when they are ignored.