Regression Models - Transmission and MPG

Ryan Wissman

Friday, March 20, 2015

Executive Summary

Fuel efficienty and transmission type are both very important and greatly debated factors when selecting a new car. This report examines the relationship between tranmission type and fuel economy to determine if there is any MPG benefit to purchasing a car of either type transmission. The data used in this report is from the 1974 Motor Trend US magazine.

Exploring the Data

First the mtcars data is loaded and some brief exploratory statistics are discovered.

```
data(mtcars); attach(mtcars)
head(mtcars,1) #Examine how the data is structured in mtcars
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4  21   6  160 110  3.9 2.62 16.46  0  1    4    4
```

```
mean_a <- mean(mpg[am=="0"]) #Automatic Transmission mean
mean_m <- mean(mpg[am=="1"]) #Manual Transmission mean
```

Quickly examining the data to determine the means we find that the average fuel economy among automatic cars is 17.14 MPG whereas the average among manual transmissions is 24.39 MPG. Furthermore, according to a boxplot of the data (*see* **Figure 1**) we could guess that the fuel efficiency of a manual transmission is greater than that of an automatic transmission. The average and median is MPG for manual transmission is distinctly higher than that of automatic transmissions. However, we cannot yet make a conclusion on based on this chart alone. First we will need to determine if a relationship does exist by using regression.

Regression Models

First we try a linear model using mpg as the outcome and transmission type (variable am, "0" denotes automatic whereas "1" denotes manual) as the predictor.

Model: Transmission Type Only

```
model_am <- lm(mpg ~ am)
summary(model_am)$adj.r.squared
```

```
## [1] 0.3384589
```

Using this model we can determine that this model can only explain about 34% (Adjusted R-squared value of 0.3598) of the variance in MPG. Therefore we should try another model to examine the other variables in the mtcars dataset.

Model: All Variables

```
model_full <- lm(mpg ~ ., mtcars)
summary(model_full)$adj.r.squared
```

```
## [1] 0.8066423
```

Fitting all the variables can explain about 81% (Adjusted R-Squared 0.8066) of the variance in MPG. Accoding to the matrix scatterplot of all the variables in mtcars (*see* **Figure 2**) the there are a number of other variables that show significant correlation with MPG. Therefore exparimentation is necessary to create a model with the highest multiple R-squared value. Furthermore, looking at our coefficnents, the variables with the lowest P value are horsepower (hp), weight (wt), 1/4 mile time (qsec), and transmission (am).

Model: Best

We will step through the itterations of variables to determine which model would be the best.

```
model_best <- step(model_full, direction="both")
```

```
## Start:  AIC=70.9
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - cyl   1    0.0799 147.57 68.915
## - vs    1    0.1601 147.66 68.932
## - carb  1    0.4067 147.90 68.986
## - gear  1    1.3531 148.85 69.190
## - drat  1    1.6270 149.12 69.249
## - disp  1    3.9167 151.41 69.736
## - hp    1    6.8399 154.33 70.348
## - qsec  1    8.8641 156.36 70.765
## <none>              147.49 70.898
## - am    1   10.5467 158.04 71.108
## - wt    1   27.0144 174.51 74.280
##
## Step:  AIC=68.92
## mpg ~ disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - vs    1    0.2685 147.84 66.973
## - carb  1    0.5201 148.09 67.028
## - gear  1    1.8211 149.40 67.308
## - drat  1    1.9826 149.56 67.342
## - disp  1    3.9009 151.47 67.750
## - hp    1    7.3632 154.94 68.473
## <none>              147.57 68.915
## - qsec  1   10.0933 157.67 69.032
## - am    1   11.8359 159.41 69.384
## + cyl   1    0.0799 147.49 70.898
## - wt    1   27.0280 174.60 72.297
##
```

```
## Step:  AIC=66.97
## mpg ~ disp + hp + drat + wt + qsec + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - carb  1    0.6855 148.53 65.121
## - gear  1    2.1437 149.99 65.434
## - drat  1    2.2139 150.06 65.449
## - disp  1    3.6467 151.49 65.753
## - hp    1    7.1060 154.95 66.475
## <none>               147.84 66.973
## - am    1   11.5694 159.41 67.384
## - qsec  1   15.6830 163.53 68.200
## + vs    1    0.2685 147.57 68.915
## + cyl   1    0.1883 147.66 68.932
## - wt    1   27.3799 175.22 70.410
##
## Step:  AIC=65.12
## mpg ~ disp + hp + drat + wt + qsec + am + gear
##
##        Df Sum of Sq    RSS    AIC
## - gear  1     1.565 150.09 63.457
## - drat  1     1.932 150.46 63.535
## <none>               148.53 65.121
## - disp  1    10.110 158.64 65.229
## - am    1    12.323 160.85 65.672
## - hp    1    14.826 163.35 66.166
## + carb  1     0.685 147.84 66.973
## + vs    1     0.434 148.09 67.028
## + cyl   1     0.414 148.11 67.032
## - qsec  1    26.408 174.94 68.358
## - wt    1    69.127 217.66 75.350
##
## Step:  AIC=63.46
## mpg ~ disp + hp + drat + wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## - drat  1     3.345 153.44 62.162
## - disp  1     8.545 158.64 63.229
## <none>               150.09 63.457
## - hp    1    13.285 163.38 64.171
## + gear  1     1.565 148.53 65.121
## + cyl   1     1.003 149.09 65.242
## + vs    1     0.645 149.45 65.319
## + carb  1     0.107 149.99 65.434
## - am    1    20.036 170.13 65.466
## - qsec  1    25.574 175.67 66.491
## - wt    1    67.572 217.66 73.351
##
## Step:  AIC=62.16
## mpg ~ disp + hp + wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## - disp  1     6.629 160.07 61.515
## <none>               153.44 62.162
```

```
## - hp    1    12.572 166.01 62.682
## + drat  1     3.345 150.09 63.457
## + gear  1     2.977 150.46 63.535
## + cyl   1     2.447 150.99 63.648
## + vs    1     1.121 152.32 63.927
## + carb  1     0.011 153.43 64.160
## - qsec  1    26.470 179.91 65.255
## - am    1    32.198 185.63 66.258
## - wt    1    69.043 222.48 72.051
##
## Step:  AIC=61.52
## mpg ~ hp + wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## - hp    1     9.219 169.29 61.307
## <none>              160.07 61.515
## + disp  1     6.629 153.44 62.162
## + carb  1     3.227 156.84 62.864
## + drat  1     1.428 158.64 63.229
## - qsec  1    20.225 180.29 63.323
## + cyl   1     0.249 159.82 63.465
## + vs    1     0.249 159.82 63.466
## + gear  1     0.171 159.90 63.481
## - am    1    25.993 186.06 64.331
## - wt    1    78.494 238.56 72.284
##
## Step:  AIC=61.31
## mpg ~ wt + qsec + am
##
##        Df Sum of Sq    RSS    AIC
## <none>              169.29 61.307
## + hp    1     9.219 160.07 61.515
## + carb  1     8.036 161.25 61.751
## + disp  1     3.276 166.01 62.682
## + cyl   1     1.501 167.78 63.022
## + drat  1     1.400 167.89 63.042
## + gear  1     0.123 169.16 63.284
## + vs    1     0.000 169.29 63.307
## - am    1    26.178 195.46 63.908
## - qsec  1   109.034 278.32 75.217
## - wt    1   183.347 352.63 82.790
```

```
summary(model_best)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Appendix

Data and figures that accompany the report.

Figure 1: Boxplot summarizing both automatic and manual transmission types relative to MPG

```
boxplot(mpg ~ am, xlab="Transmission Type", ylab="Miles per Gallon (MPG)", xaxt="n",
        main="Transmission and MPG",
        col=c("green", "red"))
axis(1, at=1:2, labels=c("Automatic", "Manual"))
```

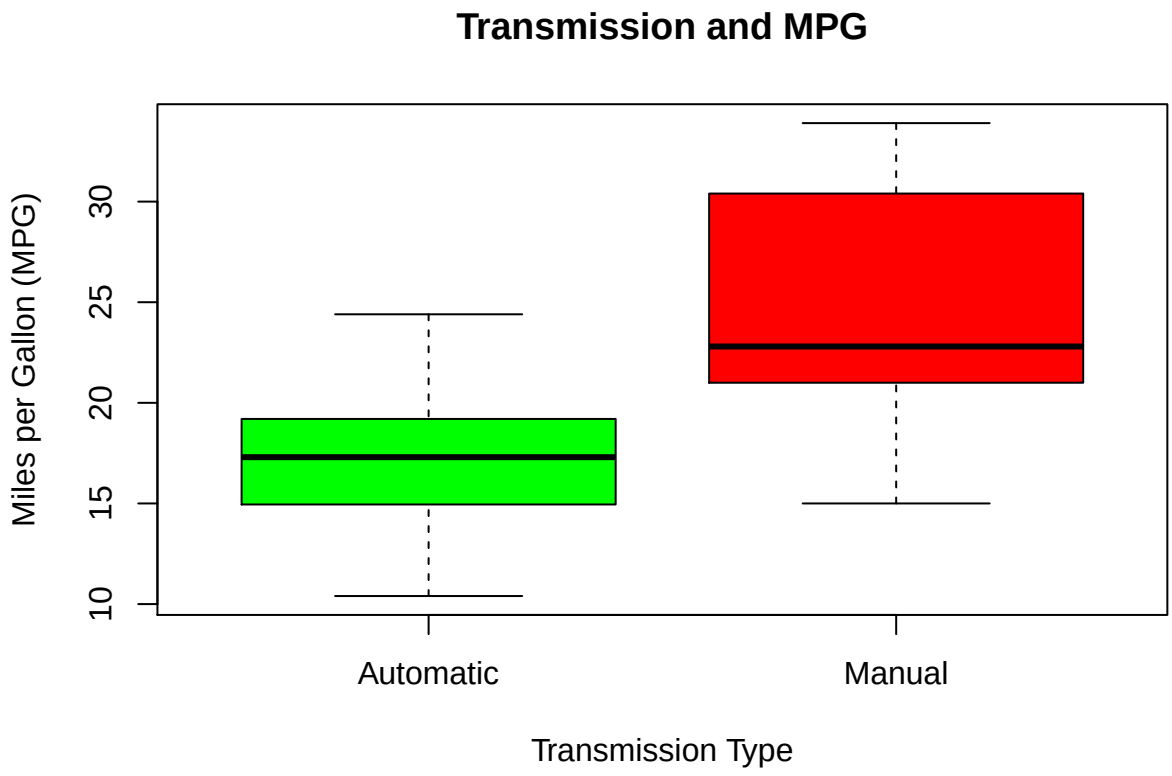

Figure 2:

Matrix scatterplot covering the variances between many different variables in the mtcars dataset.

```
pairs(mtcars, main="Matrix Scatterplot of all Variables", panel=function(x,y){
  points(x,y)
  abline(lm(y~x), col="red")
})
```