

Regression Models Course 7 Project

Ed Loessi

March 24, 2018

JHU Regression Models Course 7 - Peer Graded Project

Introduction - Executive Summary

In this project I am utilizing data from the 1974 Motor Trend Magazine Study covering 32 vehicles. The purpose of this particular analysis is to determine which type of transmission is better for higher gas mileage, an automatic transmission, or a manual transmission, and to build a best fit model comparing the two transmissions and other important variables. In completing this analysis, I performed basic exploratory, tested a few linear models, and utilized a stepwise model to determine the best combination of variables to determine a vehicle's MPG. The final analysis showed that the MPG for a vehicle is in fact higher when equipped with a manual transmission and that a model that contains the weight of the vehicle, its quarter-mile race time, and its transmission type will allow you to explain approximately 85% using the Multiple R-squared.

Load the initial dataset

The first step is to load the dataset and create factors for the vs and am variables for future processing.

```
data(mtcars)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
```

Summarize and review the data

The next step is to look through the number of observations of which there are 32 and the number of variables of which there are 11. Additionally I ran a pairs comparison to test the connections that the variables show between each other. Note - the Appendix (Figure 1) contains the pairs grid.

Observations and Variables

You can see the breakout of the variables as well as a number of the observations.

```
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  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
## $ am  : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
## $ gear: num   4  4  4  3  3  3  4  4  4 ...
## $ carb: num   4  4  1  1  2  1  4  2  2  4 ...
```

Key statistical summary

The summary can be reviewed to better understand some of the key statistical data.

```
summary(mtcars)
```

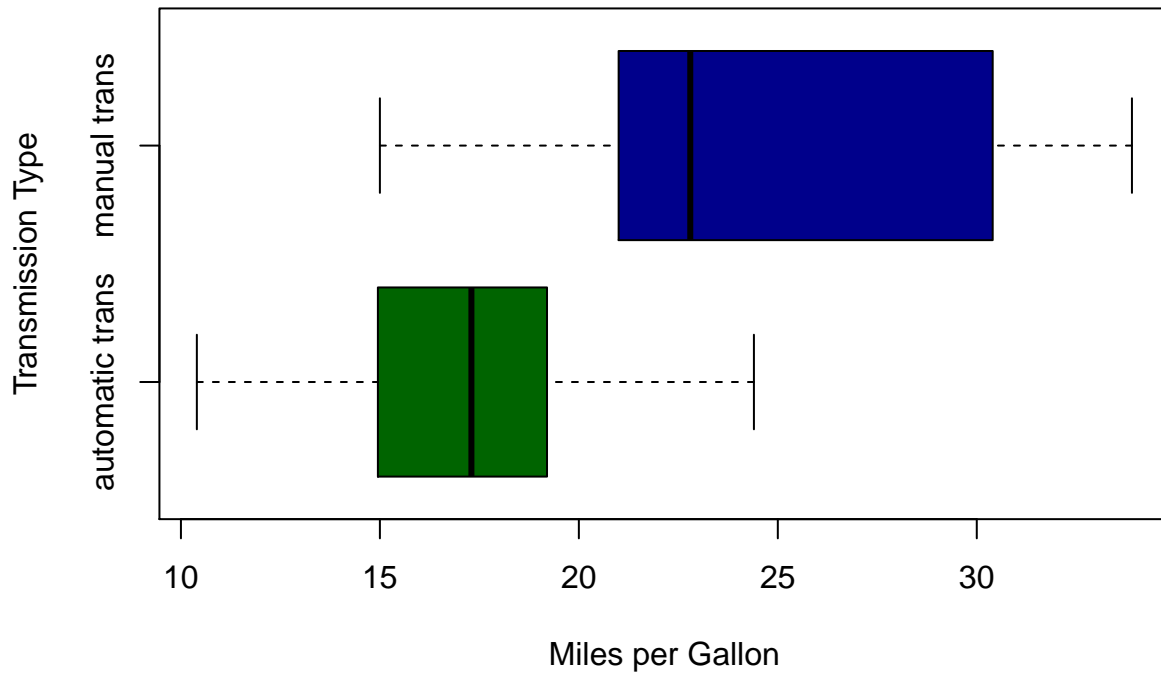
```
##           mpg           cyl           disp           hp
##  Min.      :10.40   Min.      :4.000   Min.      : 71.1   Min.      : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean    :20.09   Mean    :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           am
##  Min.      :2.760   Min.      :1.513   Min.      :14.50   0:18   0:19
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1:14   1:13
##  Median :3.695   Median :3.325   Median :17.71
##  Mean    :3.597   Mean    :3.217   Mean    :17.85
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90
##  Max.    :4.930   Max.    :5.424   Max.    :22.90
##           gear           carb
##  Min.      :3.000   Min.      :1.000
##  1st Qu.:3.000   1st Qu.:2.000
##  Median :4.000   Median :2.000
##  Mean    :3.688   Mean    :2.812
##  3rd Qu.:4.000   3rd Qu.:4.000
##  Max.    :5.000   Max.    :8.000
```

Automatic vs Manual Transmission Plotting

I ran an initial boxplot to see what the differences were between a manual and automatic transmission regarding MPG. The boxplot shows that the manual transmission performs better in terms of having a higher MPG.

```
boxplot(mpg ~ am, data = mtcars,
        col = c("dark green", "dark blue"),
        xlab = "Miles per Gallon",
        ylab = "Transmission Type",
        main = "Miles Per Gallon by Type of Transmission",
        names = c("automatic trans", "manual trans"),
        horizontal = T)
```

Miles Per Gallon by Type of Transmission



T-test for the automatic vs. manual transmissions

The t-test below renders a p-value of 0.001374 which is < 0.05 the standard marker for significance meaning that there is a difference between the two transmissions when measured against the dependent variable of MPG.

```
auto=subset(mtcars,select=mpg,am==0)
manual=subset(mtcars,select=mpg,am==1)
t.test(auto>manual)
```

```
##
##  Welch Two Sample t-test
##
## data:  auto and manual
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

Trial a series of linear regression models

In the next several steps I used a simple regression model to understand the mpg and transmission relationship, then a multivariate model to understand all of the variables in relation to the mpg, then a stepwise regression to choose the best variables to combine to determine mpg.

Simple Regression model

The simple regression model below shows that the manual transmission would be expected to outperform the automatic transmission by 7.24 miles per gallon given no other variables to consider and that it would explain 36% of the variance.

```
regSIM <- lm(mpg~am,mtcars)
summary(regSIM)

##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## 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 ***
## am1           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
```

Multivariate model

The multivariate model takes into account all variables. However in this model the manual transmission only outperforms the automatic transmission by 2.5 miles per gallon. This model explains 86% of the variance. However, in this model many of the variables are not significant.

```
regTOT <- lm(mpg~.,mtcars)
summary(regTOT)

##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.30337    18.71788   0.657   0.5181
```

```
## cyl          -0.11144    1.04502   -0.107    0.9161
## disp         0.01334    0.01786    0.747    0.4635
## hp          -0.02148    0.02177   -0.987    0.3350
## drat         0.78711    1.63537    0.481    0.6353
## wt          -3.71530    1.89441   -1.961    0.0633 .
## qsec         0.82104    0.73084    1.123    0.2739
## vs1         0.31776    2.10451    0.151    0.8814
## am1         2.52023    2.05665    1.225    0.2340
## gear         0.65541    1.49326    0.439    0.6652
## carb        -0.19942    0.82875   -0.241    0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869, Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Stepwise regression

In order to find the best set of variables I used the stepwise model below. The variables of weight, quarter-mile race time, and transmission type when combined provide the strongest model while explaining 85% of the variance. In this model, the manual transmission outperforms the automatic transmission by 2.93 miles per gallon.

```
regSR=step(regTOT,trace=0)
summary(regSR)
```

```
##
## 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 ***
## am1         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
```

Final Analysis

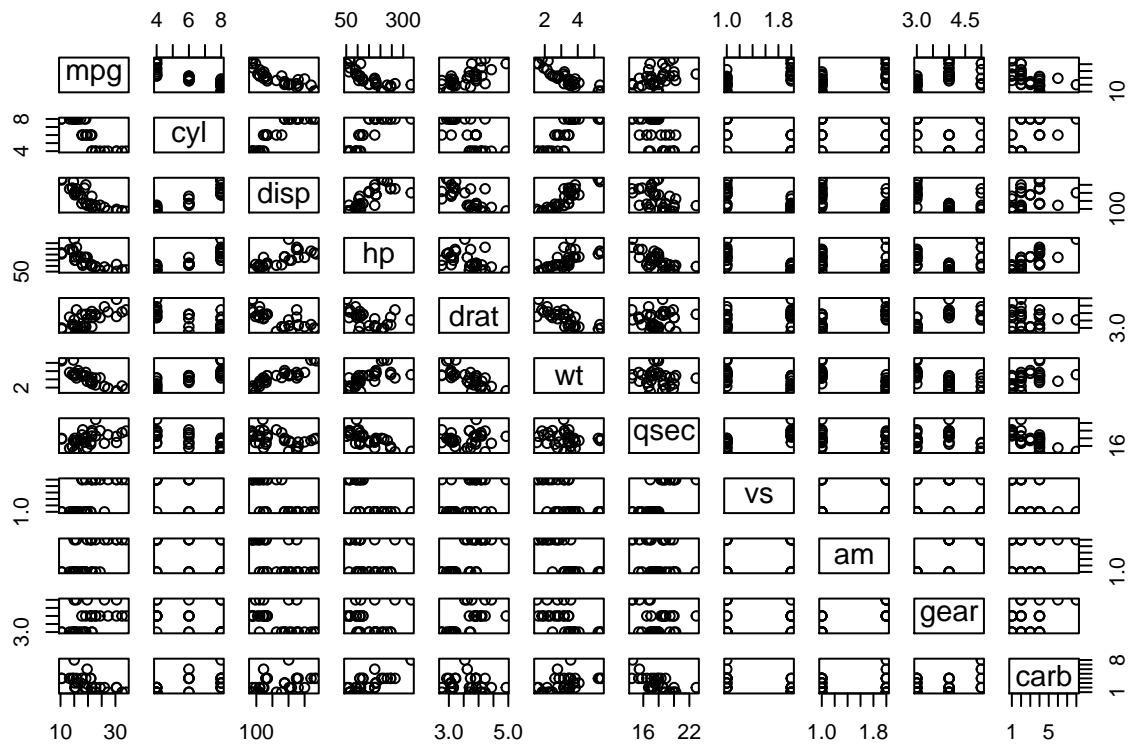
The final model above shows that weight, quarter-mile race time, and transmission type are all statistically significant regarding the MPG. This model shows that the MPG when all other factors are held constant will

improve by 2.93 miles per gallon over the automatic transmission, which answers the original question of which transmission is better for a higher MPG (manual transmission).

Appendix

Graph 1 - Pairs Analysis

```
pairs(mtcars)
```



Graph 2 - Residuals Analysis

```
plot(regSR, which=c(1:1))
```

