Regression Models Project - Motor Trend Data Analysis

by Sanjib Pradhan July 2016

Executive Summary

Motor Trend, an automobile trend magazine is interested in exploring the relationship between a set of variables and miles per gallon (MPG) outcome. In this project, we will analyze the mtcars dataset from the 1974 Motor Trend US magazine to answer the following questions:

-Is an automatic or manual transmission better for miles per gallon (MPG)?

-How different is the MPG between automatic and manual transmissions?

Using simple linear regression analysis, we determine that there is a significant difference between the mean MPG for automatic and manual transmission cars. Manual transmissions achieve a higher value of MPG compared to automatic transmission. This increase is approximately 1.8 MPG when switching from an automatic transmission to a manual one, with all else held constant.

Model Selection

After experimentation, the variables cyl, hp and wt were chosen as variables most associated with the the independent variable selected, am.

```
library(ggplot2)
library(lattice)
data(mtcars)
mtcars[1:3, ] # Sample Data
```

```
##                 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
```

```
dim(mtcars)
```

```
## [1] 32 11
```

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$am <- factor(mtcars$am, labels = c("Automatic","Manual"))

basemodel <- lm(mpg ~ am, data = mtcars)
model <- lm(mpg ~ cyl + hp + wt + am, data = mtcars)
summary(basemodel)
```

```
## 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 ***
##  amManual       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
```

Now below is the summary of the multi variable model.

```
summary(model)
```

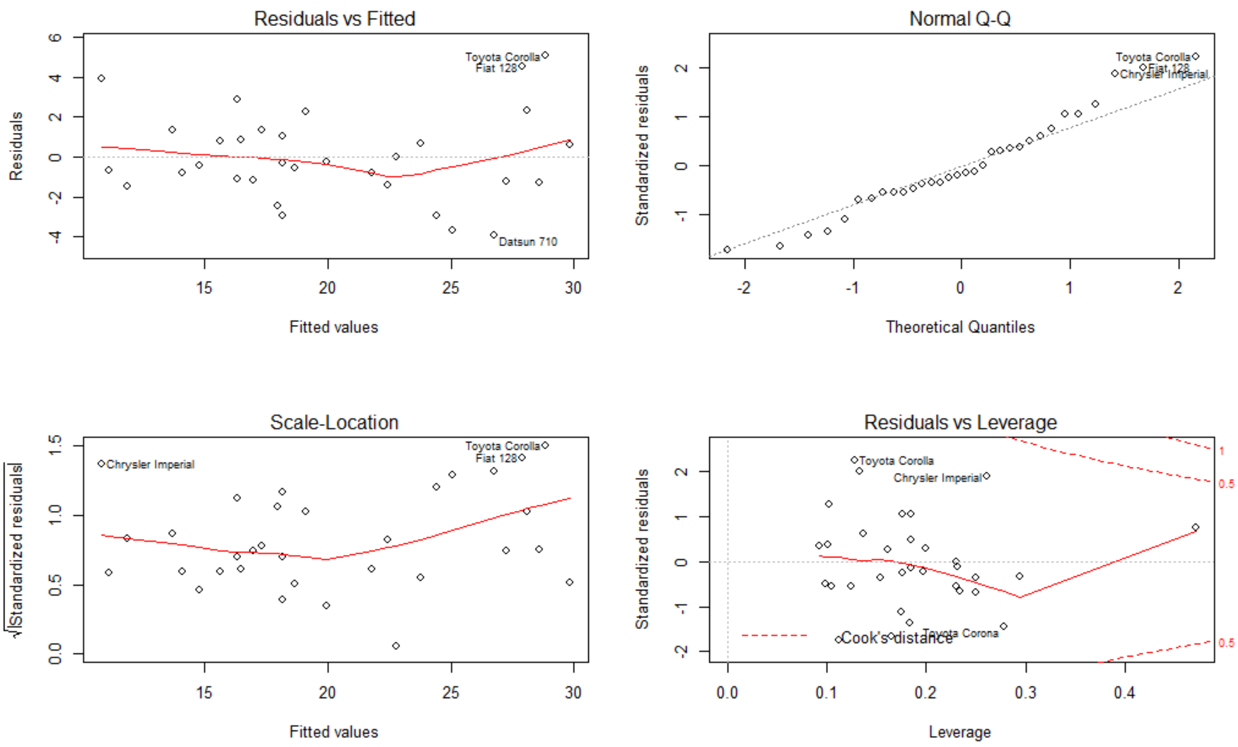
```
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
##  (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
##---
## Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659,    Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

When we examine the adjusted R-squared value, the base model has an adjusted R-squared value of 0.3385, compared to the 0.8401 of the combined model. We can say that approximately 84% of the variability is explained by the combined model above.

Model Residual and Diagnostics

In this section, we have the residual plots of our regression model along with computation of regression diagnostics for our liner model. This exercise helped us in examining the residuals and finding leverage points to find any potential problems with the model.

```
par(mfrow = c(2, 2))
plot(model)
```



From the above graphs, we can make a few observations about the combined model. The randomness of the distribution of the points in the Residuals vs. Fitted graph confirms the variable independence. The linearity of the Normal Q-Q graph indicates that the residuals are distributed under a normal distribution. The labelled points appear to be leverage points above the rest of the points.

Statistical Inference

In this section, we perform a t-test on the two subsets of mpg data: manual and automatic transmission assuming that the transmission data has a normal distribution and tests the null hypothesis that they come from the same distribution. Based on the t-test results, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.

```
t.test(mpg ~ am, data = mtcars)
```

```
##      Welch Two Sample t-test
##
##  data:  mpg by am
##  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 in group Automatic    mean in group Manual
##               17.14737                24.39231
```

From these results, we can reject the null hypothesis saying that the effect on mpg of manual and automatic transmissions are the same

Conclusion

Based on the analysis done in this project, we can conclude that:

- Cars with Manual transmission get 1.8 more miles per gallon compared to cars with Automatic transmission. (1.8 adjusted for hp, cyl, and wt).
- mpg will decrease by 2.5 for every 1000 lb increase in wt.
- mpg decreases negligibly (only 0.32) with every increase of 10 in hp.
- If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 3 and 2.2 respectively (adjusted by hp, wt, and am).

This model has the Residual standard error as 2.833 on 15 degrees of freedom. And the Adjusted R-squared value is 0.779, which means that the model can explain about 78% of the variance of the MPG variable. However, none of the coefficients are significant at 0.05 significant level.

Appendix

Figure 1 - Pairs plot for the “mtcars” dataset

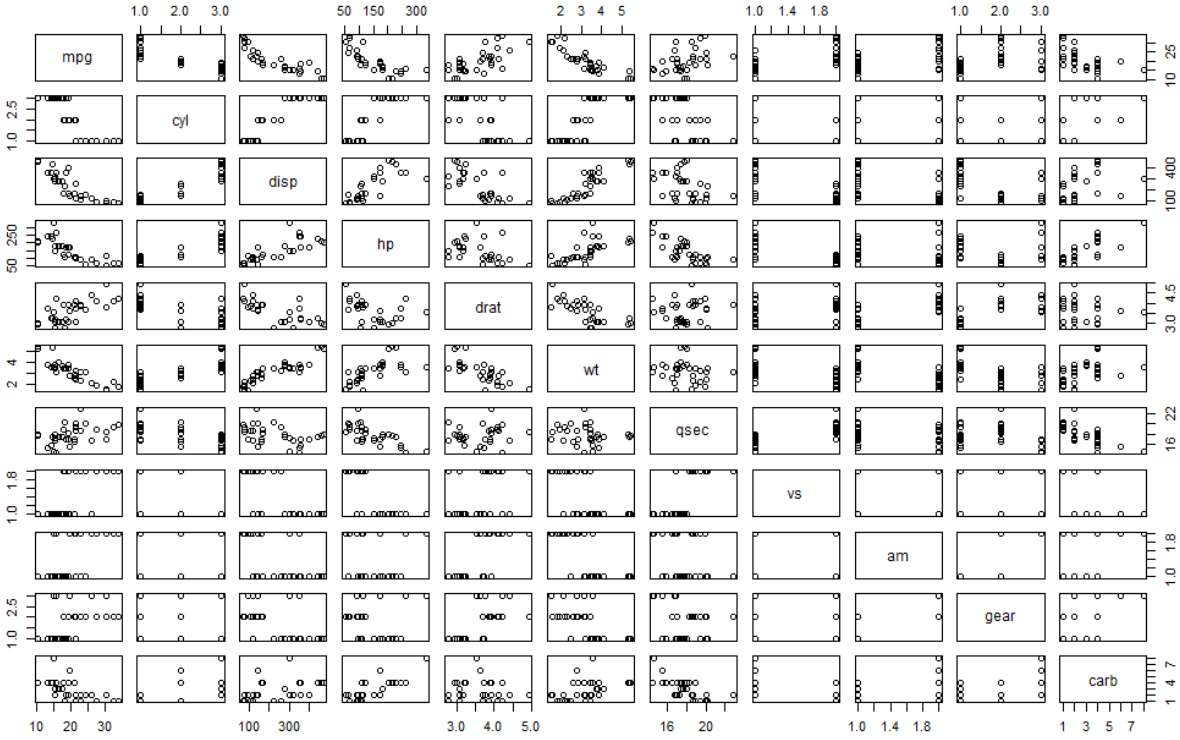
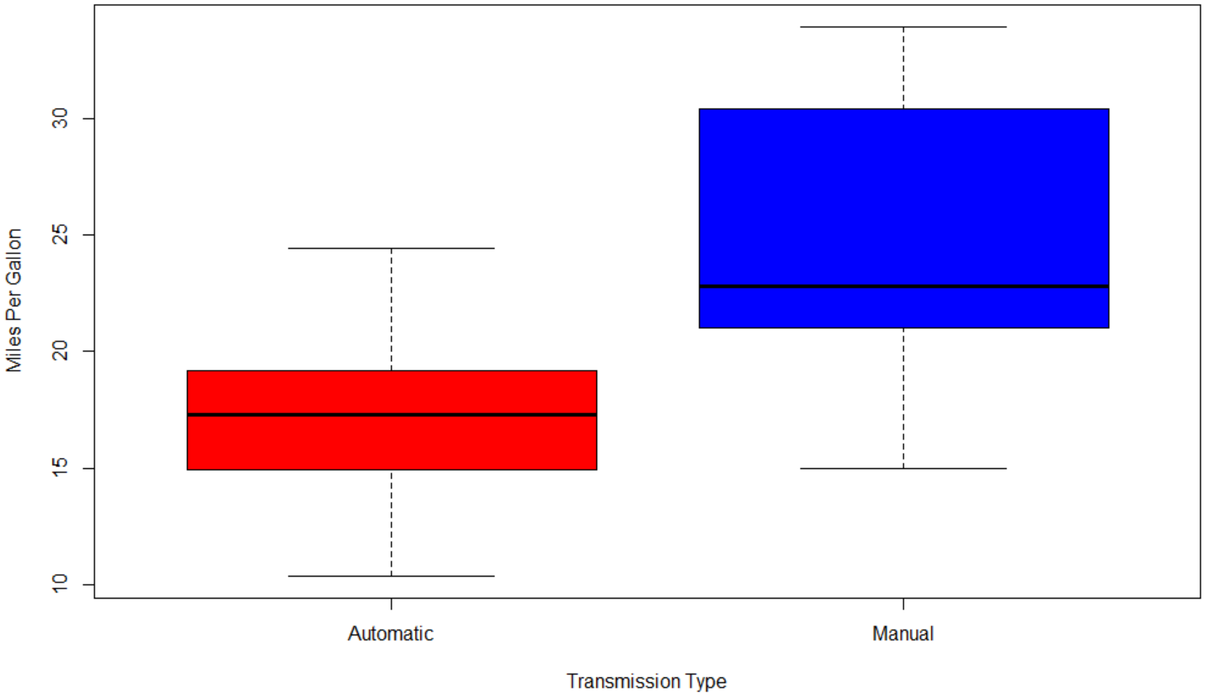


Figure 2 - Boxplot of miles per gallon by transmission type


Miles Per Gallon
Automatic
Manual
Transmission Type