Automatic or Manual Transmission - Which Has Better MPG?

Samir Yelne

March 27, 2016

Executive Summary

This project explores the relationship between miles-per-gallon (MPG) and other variables in the mtcars data set. In particular, the analysis attempts to determine whether an automatic or manual transmission is better for MPG. Therefore, 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?

Data Processing

We begin by loading in the mtcars dataset and transforming certain variables into factors.

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

Exploratory Analysis

We explore the relationships between different variables. We came to know variables cyl,disp,hp,wt are negatively correlated with mpg while variables drat,vs are positively correlated. (Appendix)

We also look at the effects of transmission type on mpg. The plot clearly shows that the manual transmission tends to have higher mpg as compared to automatic transmission. (Appendix)

Regression Analysis

Model building and selection

At the beginning we create a linear regression model with all variables as predictors and mpg as outcome variable. Then we perform stepwise model selection in order to select significant predictors for the final, best model.The step function select the best variables from them using both forward selection and backward elimination methods using AIC algorithm. This ensures that we have included useful variables while omitting ones that do not contribute significantly to predicting mpg.

```r
model1 <- lm(mpg~.,data=mtcars)
best <- step(model1,direction="both")
```

The summary of the best model is shown below.It shows the variables cyl,wt,hp as the confounders and am as the independent variable.

```r
summary(best)
```

```
##
## 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
```

The adjusted R-squared value of 0.8401 which is the maximum obtained considering all combinations of variables. From these results we can conclude that about 84% of the variability is explained by the above model.

Next we compare a model with only am as the predictor variable and the best model.

```r
model2 <- lm(mpg~am,data=mtcars)
anova(model2,best)
```

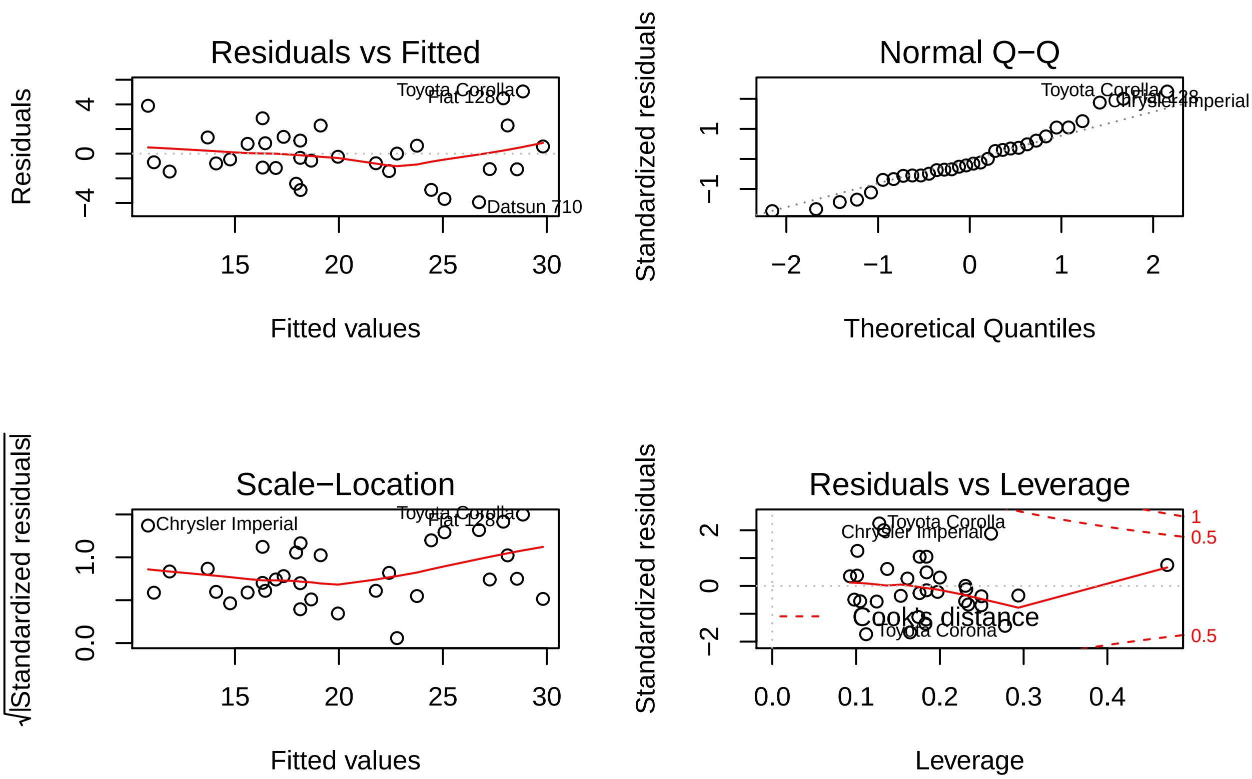
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value obtained is highly significant and we reject the null hypothesis that the confounder variables cyl, hp and wt don't contribute to the accuracy of the model.

Model Residuals and Diagnostics

In this section, we have the residual and diagnostics plots for our regression model.

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



Statistical Inference

In this section we perform a t test. Based on the t-test results, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.

```r
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
```

Conclusions

Thus, we conclude :

1. 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).

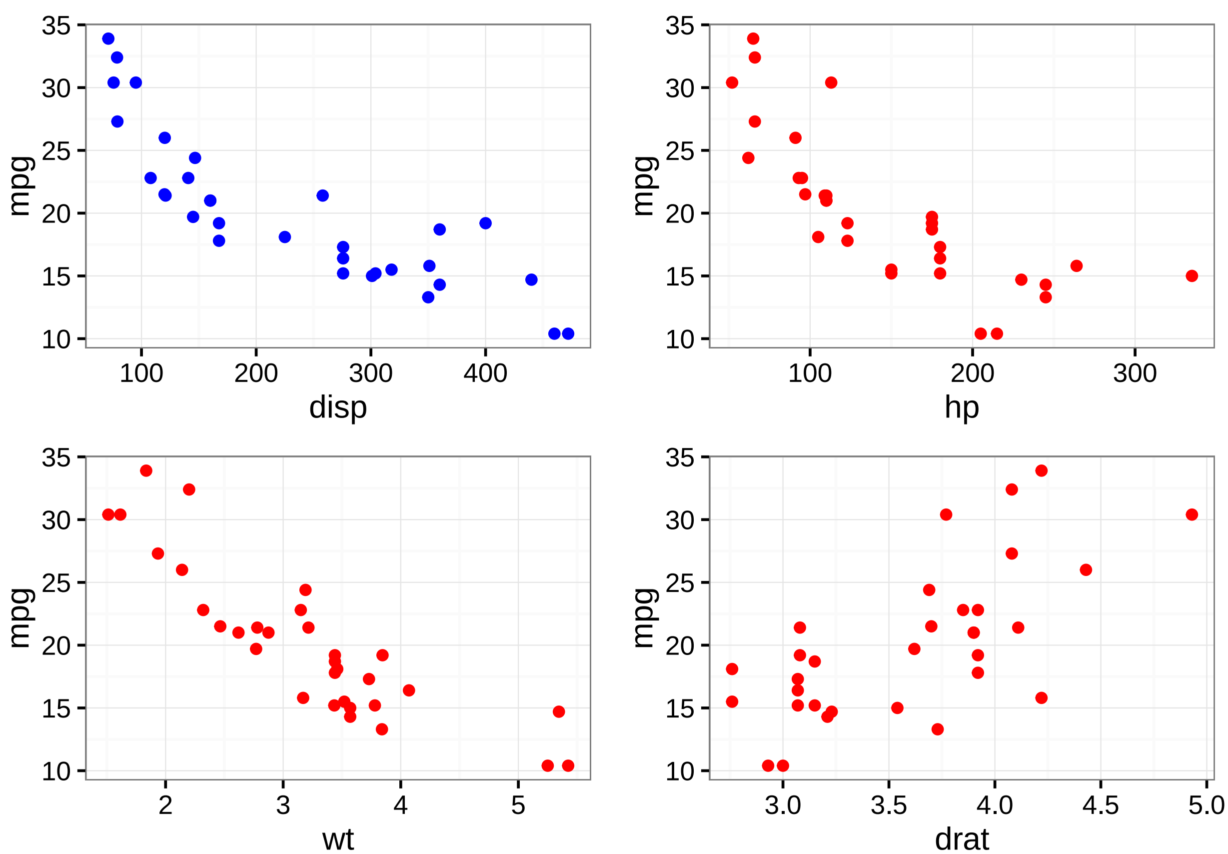
2. mpg will decrease by 2.5 for every 1000 lb increase in wt.

3. mpg decreases negligibly (only 0.0321) with every increase of 10 in hp.

4. If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 3 and 2.16 respectively (adjusted by hp, wt, and am).

Appendix

```
library(ggplot2)
library(gridExtra)
data(mtcars)
g1 <- ggplot(mtcars,aes(disp,mpg))+geom_point(color="blue")+theme_bw()
g2 <- ggplot(mtcars,aes(hp,mpg))+geom_point(color="red")+theme_bw()
g3 <- ggplot(mtcars,aes(wt,mpg))+geom_point(color="red")+theme_bw()
g4 <- ggplot(mtcars,aes(drat,mpg))+geom_point(color="red")+theme_bw()
grid.arrange(g1,g2,g3,g4,ncol=2,nrow=2)
```



```r
data(mtcars)
library(ggplot2)
mtcars$am <- factor(mtcars$am)
ggplot(mtcars,aes(am,mpg))+geom_boxplot(aes(fill=am))+theme_bw()
```