Regression Models Assignment

PJ

May 13, 2016

Executive Summary

The mtcars dataset from R was analyzed to evaluate whether there is a relationship between fuel efficiency(mpg) and transmission type which is captured as the am variable (automatic, am=0 or manual, am=1). It was found that lighter vehicles are more efficient if they are manual.

Introduction

The mtcars dataset was extracted from the 1974 Motor Trend US magazine. It comprises of the fuel efficiency data along with 10 other aspects of 32 car models produced between 1973 - 1974.

```
data(mtcars)
head(mtcars,4) #sampling the 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
## Hornet 4 Drive 21.4  6  258 110 3.08 3.215 19.44  1  0    3    1
```

Tidying the data

We shall turn some of the variables in the dataset into factors.

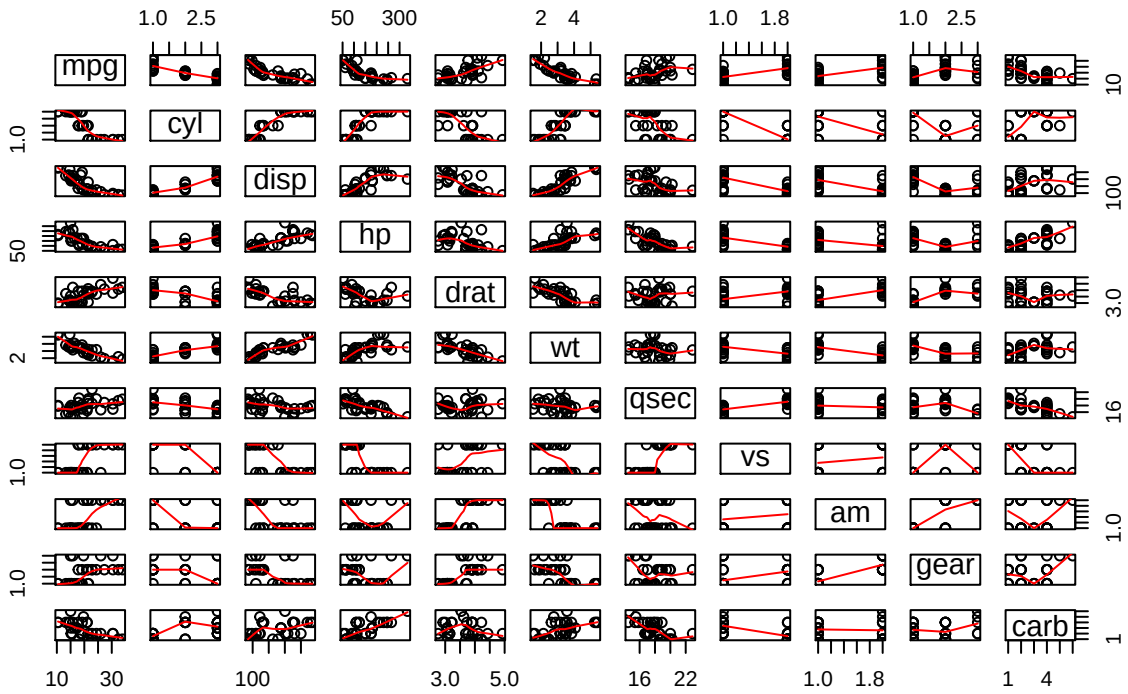
```
mtcars$gear <- as.factor(mtcars$gear)
mtcars$am <- as.factor(mtcars$am)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$carb <- as.factor(mtcars$carb)
```

Exploratory analysis

Let us look at how different variables in the dataset affect each other using a pairs graph.

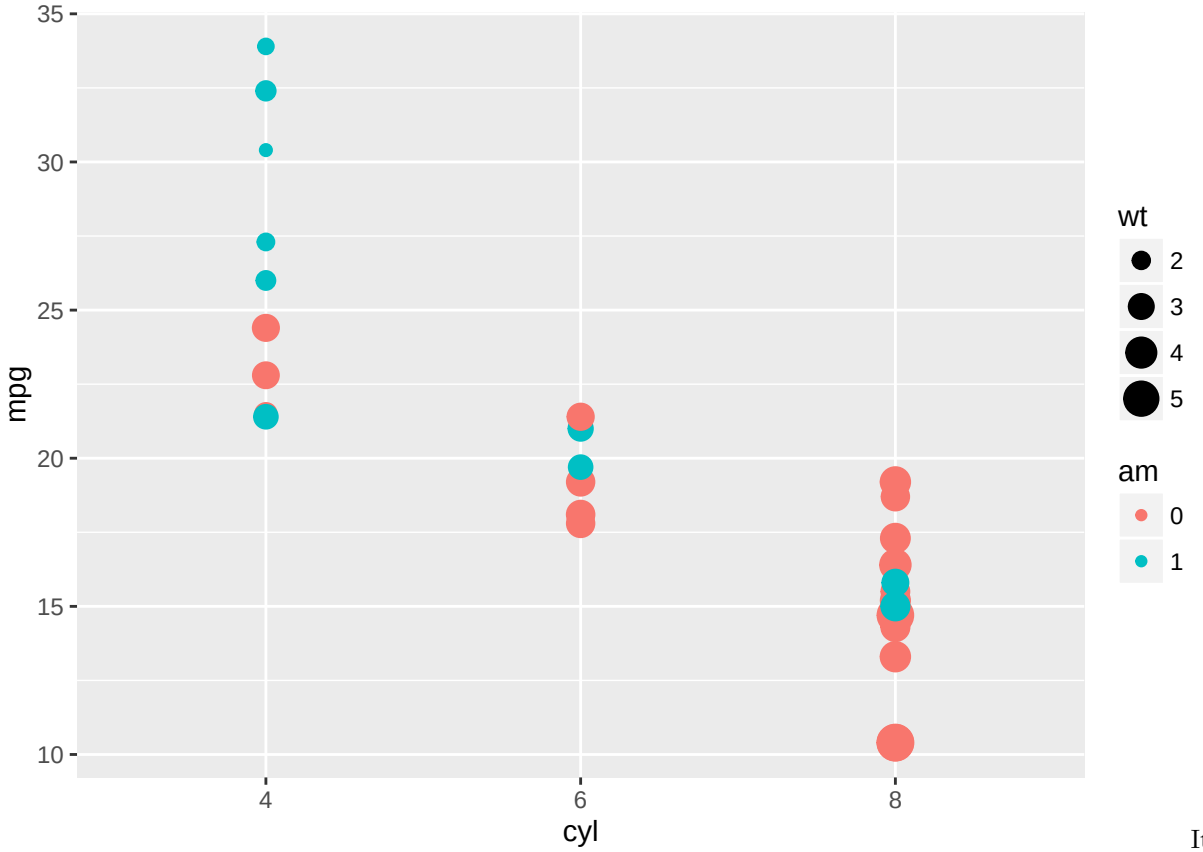
```
pairs(mtcars, panel=panel.smooth, main="Pair Graph of Motor Trend Car Data")
```

Pair Graph of Motor Trend Car Data



Then we shall look at how the mpg variable of both transmission types line up against some of the variables that would seem to affect the car's performance based on the pairs plot (like number of cylinders and weight).

```
library(ggplot2)
qplot(cyl,mpg,color = am , data = mtcars, size = wt)
```



It can be gleaned from the figure that automatic cars are generally heavier than manual cars of the same cylinder category and the mpg efficiency for manual transmission cars is generally higher in the plot. ## Regression modeling Now let us create linear models using the different variables available to us.

```
fit1 <- lm(mpg ~ am,data = mtcars)
fitsome <- lm(mpg ~am + wt + hp,data =mtcars)
fitall <- lm(mpg ~. , data=mtcars) #results hidden
```

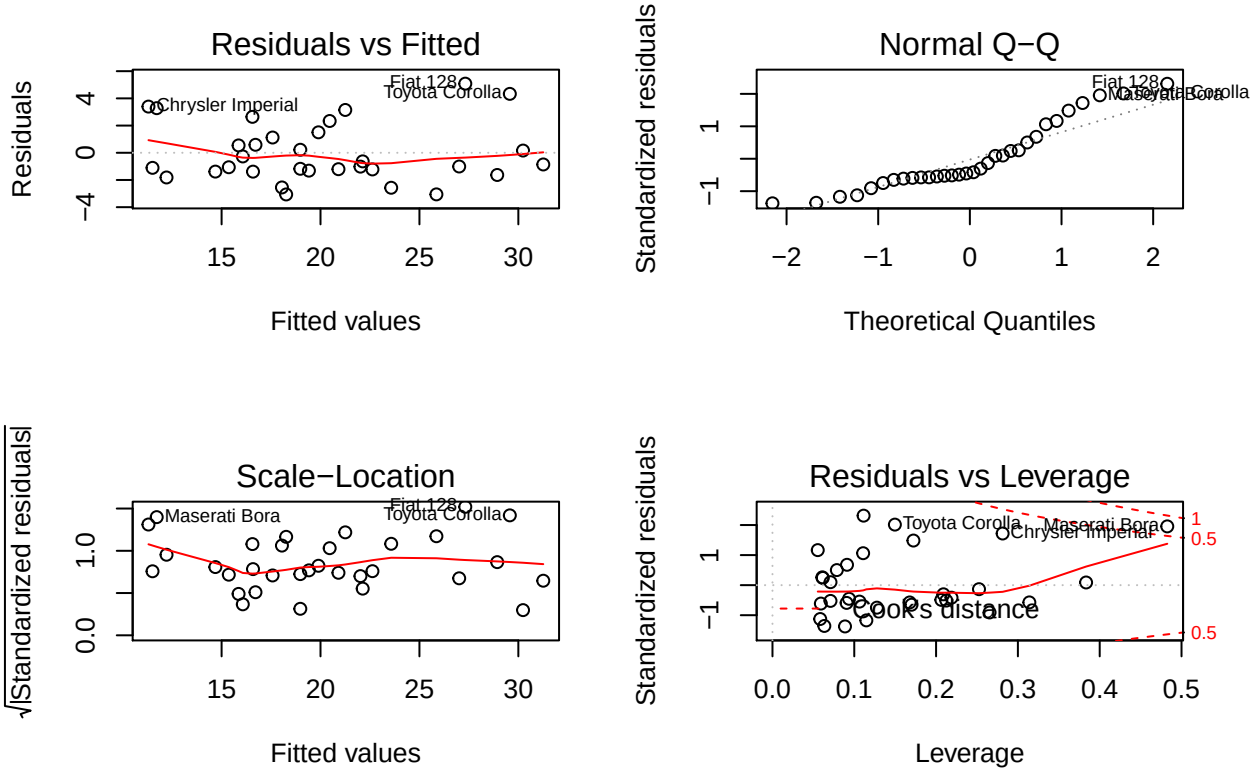
We're getting close to over 80% accuracy (adjusted r squared >0.8) by now. By accounting for the interaction between weight and transmission, we notice adjusted r squared rise to 0.85 and the P value is significant (<0.05 for all model variables)

```
fitAMwithWT <- lm(mpg ~ am + hp+ wt:am ,data =mtcars)
summary(fitAMwithWT)
```

```
##
## Call:
## lm(formula = mpg ~ am + hp + wt:am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0639 -1.3315 -0.9347  1.2180  5.0822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.947333   2.723411  11.363 8.55e-12 ***
## am1         11.554813   4.023277   2.872 0.007845 **
```

```
## hp          -0.026949   0.009796  -2.751 0.010477 *
## am0:wt      -2.515586   0.844497  -2.979 0.006052 **
## am1:wt      -6.093495   1.540277  -3.956 0.000497 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.332 on 27 degrees of freedom
## Multiple R-squared:  0.8696, Adjusted R-squared:  0.8503
## F-statistic: 45.01 on 4 and 27 DF,  p-value: 1.451e-11
```

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



Residual Analysis and Diagnostics

According to the residual plots, we can verify the following underlying assumptions:

1. The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption.
2. The Normal Q-Q plot indicates that the residuals are normally distributed because the points lie closely to the line.
3. The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed.
4. The Residuals vs. Leverage argues that no outliers are present, as almost all the values fall well within the 0.5 bands.

In conclusion, the transmission type affects the efficiency of the cars. In general, for manual cars vs automatic cars, the fuel efficiency is (-2.515) -(-6.093) mpg = 3.578 mpg greater for manual vehicles after accounting for weight.