Regression Models Peer Graded Assignment

K Ajay Rangan

03/11/2020

```
library("plyr")
```

Automatic vs Manual Transmission for better Mileage

Instructions

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

"Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

Dataset

Import the dataset

```
data(mtcars)
head(mtcars)
```

```
##                   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
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

In the dataset the am column shows what type of transmission the car has. According to the Documentation: [, 9] am Transmission (0 = automatic, 1 = manual) We require only the mpg and the am column for our analysis. Lets create a dataframe with only these variables.

```
mydata <- mtcars[c("mpg","am")]
head(mydata)
```

```
##                   mpg am
## Mazda RX4         21.0  1
## Mazda RX4 Wag     21.0  1
## Datsun 710        22.8  1
```

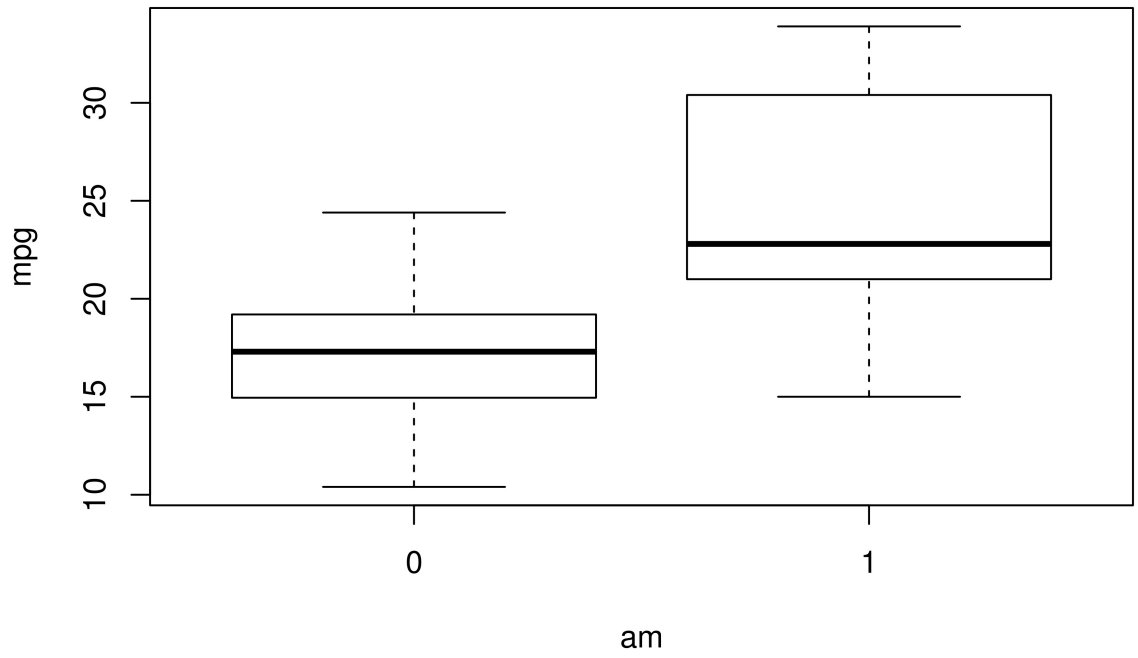
```
## Hornet 4 Drive    21.4  0
## Hornet Sportabout 18.7  0
## Valiant           18.1  0
```

Exploratory Analysis

```
summary(mydata)
```

```
##       mpg              am
##  Min.   :10.40   Min.   :0.0000
##  1st Qu.:15.43   1st Qu.:0.0000
##  Median :19.20   Median :0.0000
##  Mean   :20.09   Mean   :0.4062
##  3rd Qu.:22.80   3rd Qu.:1.0000
##  Max.   :33.90   Max.   :1.0000
```

```
boxplot(mpg~am, data = mydata)
```



A very preliminary Analysis shows the superiority of Manual Transmission in Mileage.

Lets create some models to find out, but lets also keep in mind, there may be other variables that affect this. The results of the difference in mpg aren't solely due to the type of transmission.

```r
count(mtcars, vars = "am")
```

```
##   am freq
## 1  0   19
## 2  1   13
```

There are slightly more Automatic than Manual cars in this dataset

Model

```r
premodel <- lm(mpg ~ ., mtcars);
summary(premodel)
```

```
##
## 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
## vs           0.31776    2.10451   0.151   0.8814
## am           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
```

From the table we can note that wt, am, drat, qsec, gear have high coefficients and are significant in our model

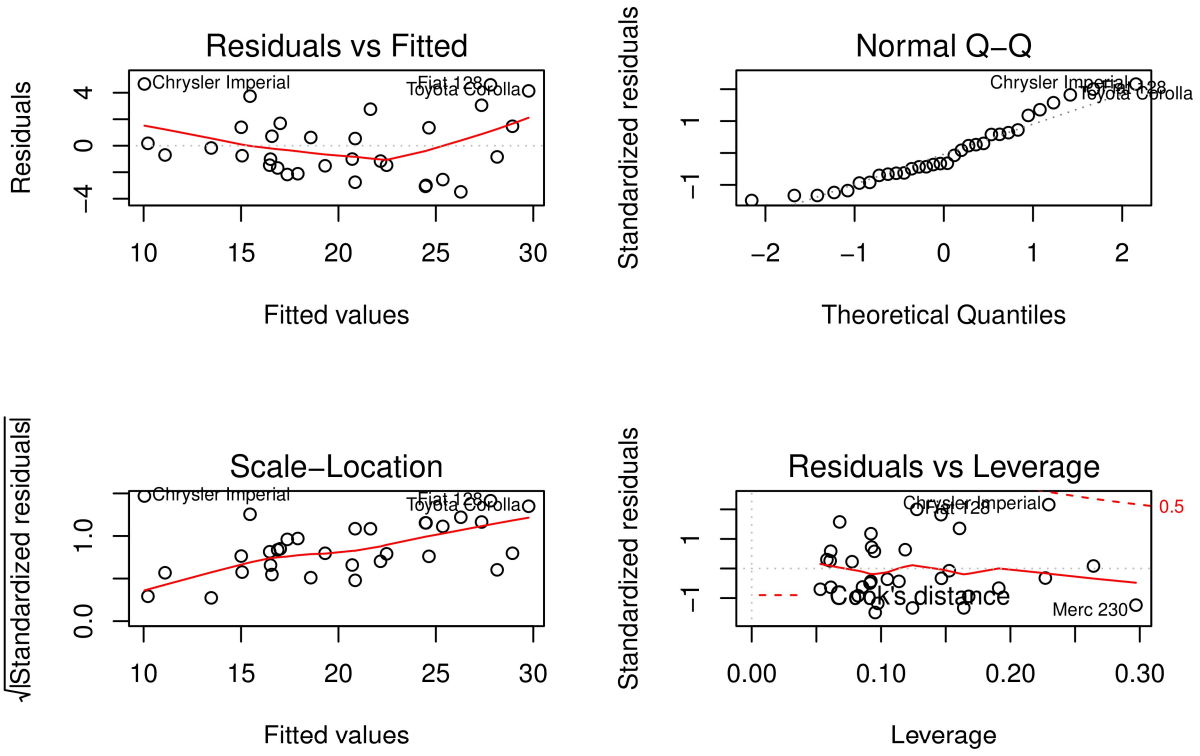
Lets compare multiple models using ANOVA

```r
anova(lm(mpg ~ am, mtcars),
      lm(mpg ~ am + wt, mtcars),
      lm(mpg ~ am + wt + qsec, mtcars),
      lm(mpg ~ am + wt + qsec + drat, mtcars),
      lm(mpg ~ am + wt + qsec + drat + gear, mtcars)
      )
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + qsec
## Model 4: mpg ~ am + wt + qsec + drat
## Model 5: mpg ~ am + wt + qsec + drat + gear
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 68.8055 8.913e-09 ***
## 3     28 169.29  1    109.03 16.9510 0.0003442 ***
## 4     27 167.89  1      1.40  0.2176 0.6447654
## 5     26 167.24  1      0.65  0.1006 0.7537011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see high P values for the 4th and 5th Model. The 3rd model seems to be a better one given lower RSS and P. This includes am, wt and qsec.

```
model2 <- lm(mpg ~ am + wt + qsec, mtcars)
par(mfrow = c(2,2)); plot(model2)
```




```
model1 <- lm(mpg~am,mtcars)
coef(model1)
```

```
## (Intercept)          am
##   17.147368    7.244939
```

```
coef(model2)
```

```
## (Intercept)          am          wt        qsec
##    9.617781    2.935837   -3.916504    1.225886
```

Executive Summary

As we can see in general Manual cars have a mileage about 7.244939 higher than automatic transmission. If we bring in other variables like wt and qsec which are significant into the picture, Manual cars have higher mileage by about 2.935837.

In either case its safe to say Manual cars have a better mileage.