Project: Regression Models

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

Preprocessing

```
# - Load dataframe
data(mtcars)

# - Analysis: Relationship between mpg and others variables
cor(mtcars$mpg,mtcars[,-1])
```

```
##            cyl       disp         hp      drat         wt     qsec
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##             vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

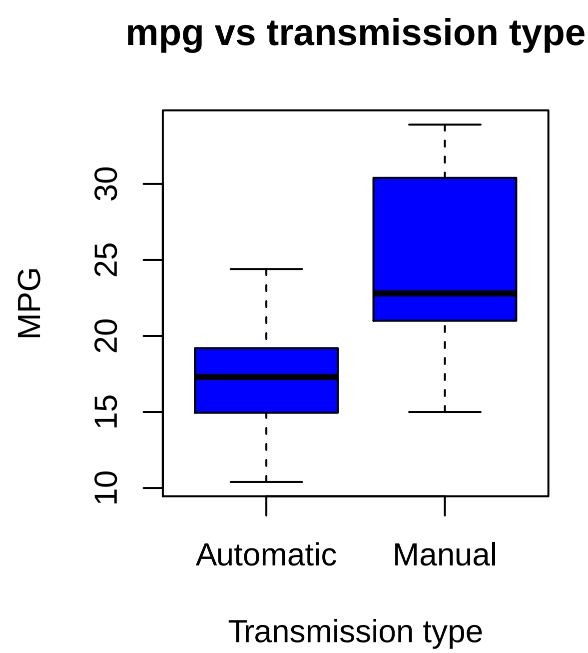
Question 1: Is an automatic or manual transmission better for MPG (miles per gallon)?

```
# - Transmission type: 0 automatic and 1 manual
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")

# - Comparison for MPG
par(mfrow = c(1, 2), pty = "s")
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars, outpch = 19, ylab="MPG",
        xlab="Transmission type",main="mpg vs transmission type", col="blue")

# - Statistical analysis
t.test(mtcars$mpg~mtcars$am,conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$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
```



The p-value is 0.001374, we may reject the null hypothesis and conclude, that automatic transmission cars have lower mpg compared with manual transmission cars

Question 2: Quantify the MPG difference between automatic and manual transmissions

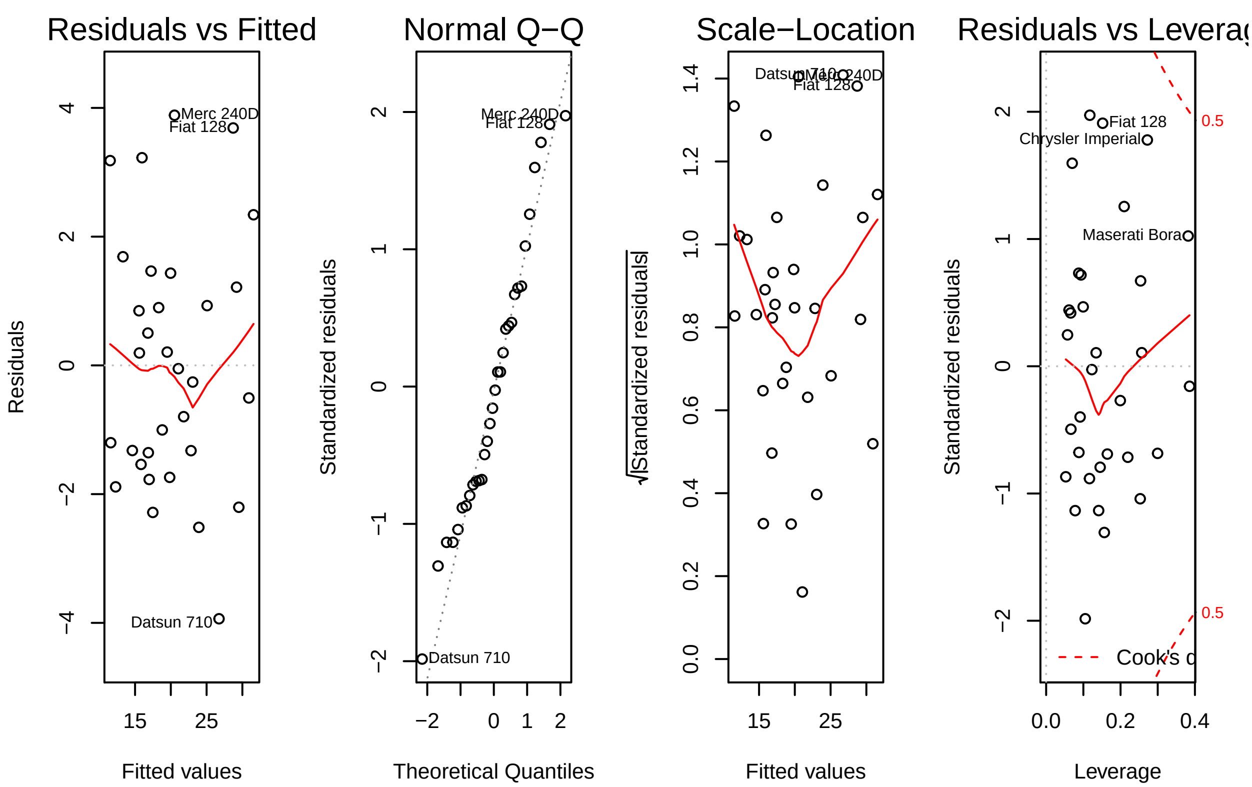
```r
# - Select a model
stepmodel <- step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
#summary(stepmodel)

# - Best Model
model <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               13.9692     5.7756   2.419  0.02259 *
## factor(am)Automatic:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
```

```
## factor(am)Manual:wt       -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)Automatic:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)Manual:qsec      1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

```r
par(mfrow=c(1,4))
plot(model)
```



Interpreting the results, we can see this model has a 89.5% total variance with an adjusted variance of 0.879

Conclusion

The mpg is largely determined by the interplay between weight, acceleration and transmission. Given the above analysis, the first question is not really answered, and should be considered in the context of weight and acceleration speed.