

Motor Trend - Regression Models

Executive Summary

This assignment will analyse `mtcars` dataset and explore the relationship between a set of variables and miles per gallon (MPG) (outcome). The purpose of this assignment is to answer the following two questions:

1. Is an automatic or manual transmission better for MPG
2. Quantify the MPG difference between automatic and manual transmissions

Data Processing

```
# Load the data
data(mtcars)
# Check the structure of the data
str(mtcars)
```

```
## 'data.frame':   32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

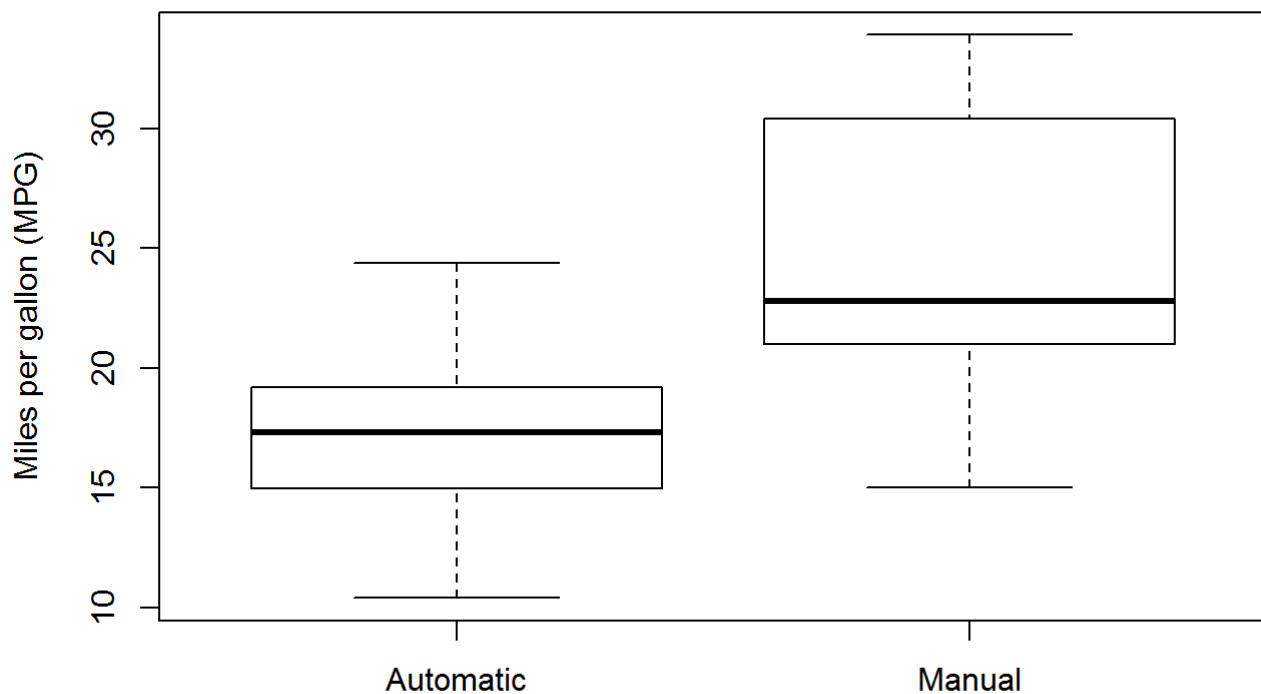
```
# Convert appropriate variables into factors
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

Data Exploring

```
head(mtcars)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs      am gear
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0   Manual    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0   Manual    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61  1   Manual    4
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1 Automatic    3
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0 Automatic    3
## Valiant        18.1   6  225 105 2.76 3.460 20.22  1 Automatic    3
##           carb
## Mazda RX4      4
## Mazda RX4 Wag  4
## Datsun 710      1
## Hornet 4 Drive  1
## Hornet Sportabout 2
## Valiant        1
```

```
# Plot variable "am" and "mpg" for visial representation
plot(mtcars$am,mtcars$mpg, ylab = "Miles per gallon (MPG)")
```



From the plot we can conclude that the Automatic and Manual transmission are very different regarding MPG. We can also find support in that conclusion by performing a `t.test` to test a null hypothesis for difference in the transmission mode in the `am` variable.

```
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
```

With p-value equal to 0.001374, we can reject the Null Hypothesis and firmly conclude that there is indeed difference b/n Automatic Transmission and Manual Transmission, regarding Miles Per Gallon.

Simple Linear Regression

```
# Perform Simple Linear Regression
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)
```

```
##
## 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
```

The Adjusted R-squared value of the linear model is 0.3385, which means that 33.8% of the response variable variation is explained by our model, and we should explore different models.

Multivariable Regression

In order to see which variables to include in our next model we can use the `stepAIC` function from MASS package to show us which are the optimal variables.

```
# step function to find variables that are optimal
library(MASS)
OptVariables <- stepAIC(lm(mpg ~., data = mtcars), direction = "both", trace = 0)
summary(OptVariables)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Adjusted R-squared = 0.8401, which means that 84% of the response variable variation is explained by our new model. That makes this model the better choice for our data.

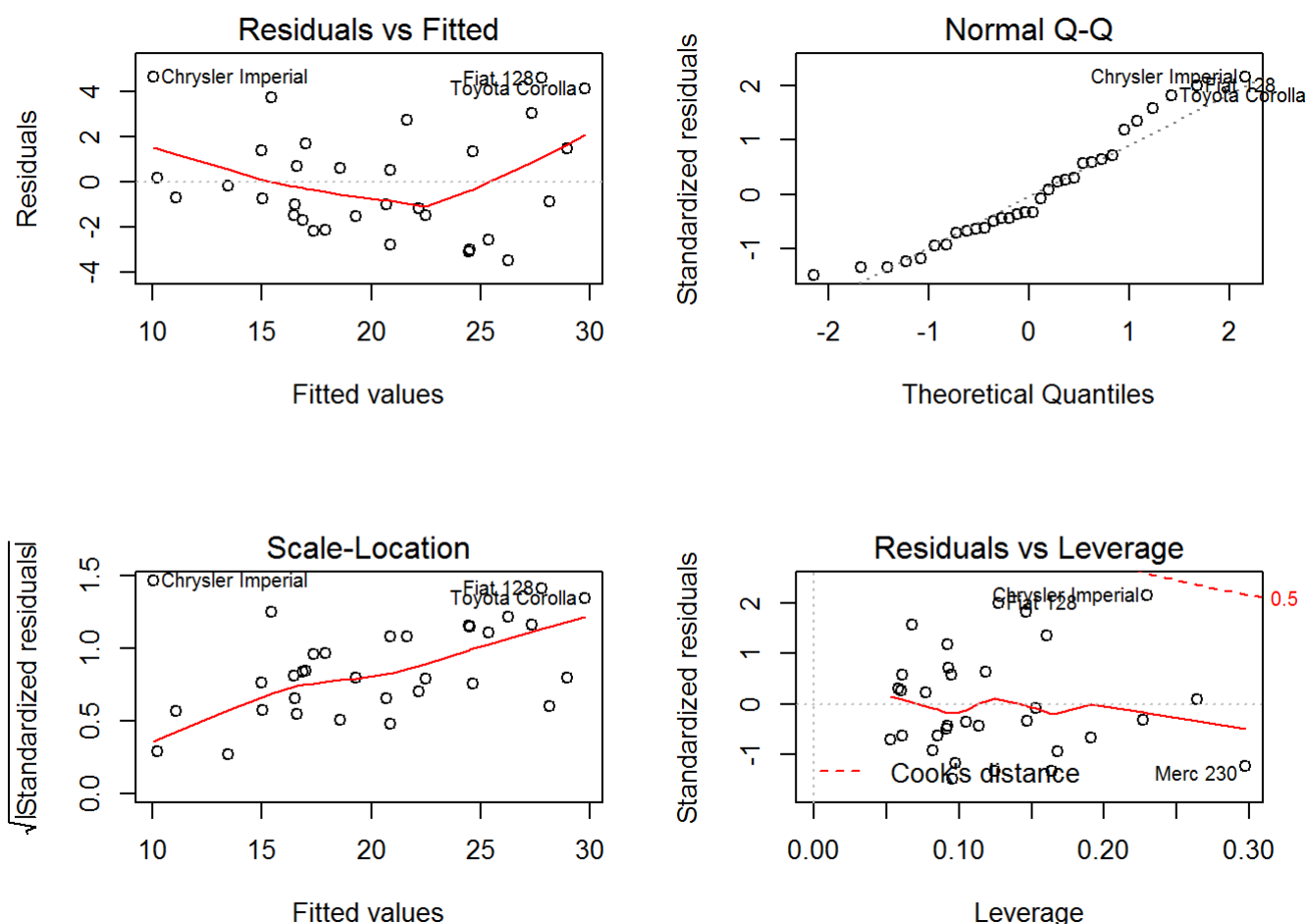
```
# Perform an ANOVA test to compare the two models
anova(fit, OptVariables)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      30 720.90
## 2      28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the p-value = 1.688e-08 we can reject the null hypothesis and to conclude that the `cyl`, `hp`, `wt` and `am` variables, all contribute to the model, which is trying to predict `miles per gallon` variable.

Diagnostic Plots

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



- Residuals vs. Fitted plot is randomly scattered - verifies the independence condition
- Normal Q-Q plot fall on the line - indicating normal distribution of the residuals
- Scale-Location plot consists of scattered points - indicating constant variance
- Cook's distance plot consist of some points - outliers or leverage points - we may investigate further.

Chrysler Imperial, Fiat 128, and Toyota Corolla are influential points

Conclusion

Answers of the two questions:

1. Is an automatic or manual transmission better for MPG?

- Manual transmission is better for *mpg*

2. Quantify the MPG difference between automatic and manual transmissions

- Manual transmission cars delivers 2.94 more *mpg* than automatic transmission cars with an adjusted *R*-squared of 0.83 and a *p*-value below 0.05