Regression Models Assignment

Executive Summary

Based on the data collected by Motor Trend magazine, this report strives to establish the relationship between Miles per Gallon and other variables. This report will specifically address the following:

1. If automatic or manual tranmission better for MPG
2. Quantification of the MPG difference between automatic and manual transmissions

It is found inconclusive to determine whether automatic or manual transmission is better as Miles per Gallon and Transmission Type is significantly influenced by weight and acceleration of the cars.

Data Processing

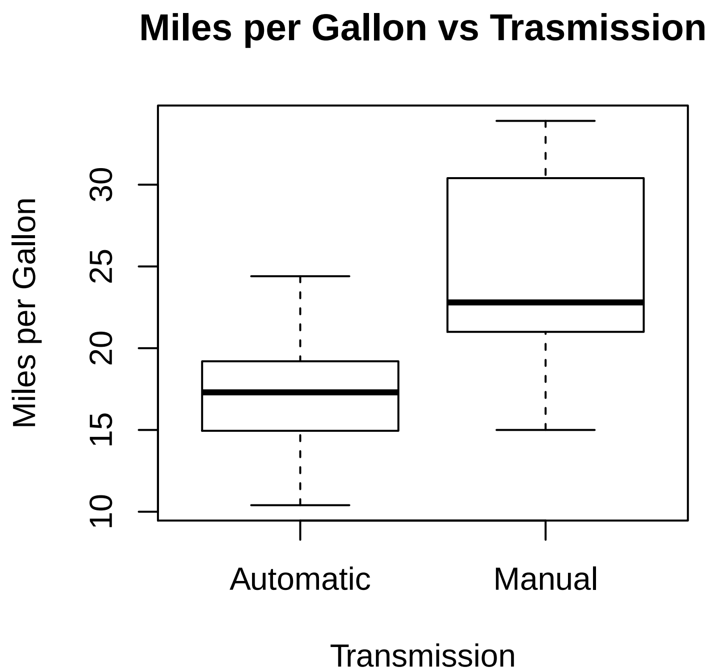
1. The data is read into R. It is determined that the data is fairly clean, therefore only requiring levels to be added for the am variable. (Please see appendix A)

```
data(mtcars);mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

Exploratory Data Analysis

1. Through a boxplot, it seems that the type of transmission affects mpg significantly.

```
boxplot(mpg~am,data=mtcars,xlab = "Transmission",ylab = "Miles per Gallon",
        main = "Miles per Gallon vs Trasmission")
```



2. To confirm the significance, a t-test is performed with the null hypothesis H0 being the case where there is no difference between automatic and manual transmission and the alternative hypothesis HA being the case where there is significant difference between automatic and manual transmission.

3. With a p-value of 0.001374, the null hypothesis is rejected.

```
test <- t.test(mpg~am,data=mtcars); test$p.value
```

```
## [1] 0.001373638
```

Regression Analysis

Initial Fit

1. Given the significance found between the two transmissions, an initial fit using linear regression is attempted (See appendix B). However the fit is not optimal as the adjusted R-Squared value is low (0.338).

Model Selection and assumption checking

1. To find the optimal sets of variables, the step() function is used.(See appendix C). The variables are am, qsec and wt.
2. The model is then checked for its assumptions and it is found to be in compliance (see Appendix D)

Proposed Model - Multivariate Regression Model

1. The model is then fitted (See Appendix E). Model1 is a better fit with an adjusted R-Squared value of 0.8336. At significance of alpha = 0.05, the model suggests that transmission type, weight and

acceleration all played significant roles. (Appendix E, coef output)

2. Further optimization is done to interactions between transmission type, weight and acceleration, resulting in model2 with adjusted R-Squared value of 0.879

```
model2 <- lm(formula = mpg ~ am:wt + am:qsec, data = mtcars)
summary(model2)$coef;summary(model2)$adj.r.squared
```

```
##                    Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)      13.9692069  5.7756116  2.418654 2.259367e-02
## amAutomatic:wt   -3.1758862  0.6362299 -4.991727 3.114029e-05
## amManual:wt      -6.0991935  0.9685466 -6.297264 9.703599e-07
## amAutomatic:qsec  0.8337859  0.2601709  3.204762 3.458031e-03
## amManual:qsec     1.4463757  0.2692125  5.372616 1.120875e-05

## [1] 0.8789714
```

Conclusion

1. Given the resulting model (model2), it is inconclusive to determine whether automatic or manual transmission is better. This is because, at significance alpha = 0.05, Miles per Gallon and Transmission Type variables are significantly influenced by weight and acceleration.

Appendix

Appendix A - Data Processing

1. Details of the dataset in shown.

```
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  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Appendix B - Initial Fit - Adjusted R-Squared

```
initial <- lm(mpg~am, data=mtcars)
summary(initial)$adj.r.squared
```

```
## [1] 0.3384589
```

Appendix C - Model Selection

```r
step(lm(data = mtcars, mpg ~ .),direction="both", trace=0)
```

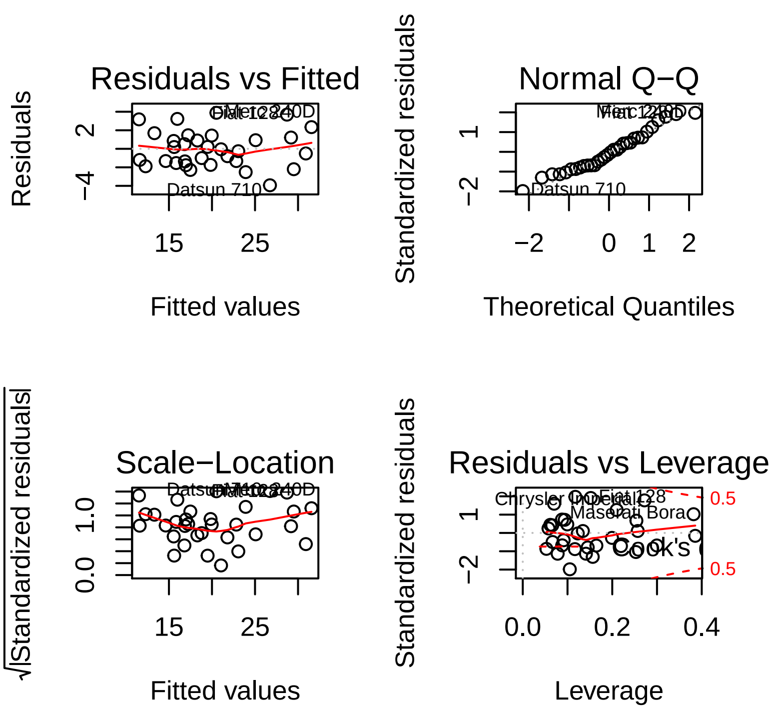
```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Coefficients:
## (Intercept)           wt         qsec     amManual
##       9.618       -3.917        1.226        2.936
```

Appendix D - Checking model assumptions

In this appendix, we shall further explore how model assumptions are checked.The four graphs below are created using R's base installation.

1. **Linearity:** From the Residuals vs Fitted graph, the values shows a linear trend, thus linear regression can be assumed.
2. **Normality:** From the Normal Q-Q plot, it can be observed that there are no outliers and values can be assumed normally distributed.
3. **Homoscedasticity:** From the Scale-Location graph, it can be seen that the values are randomly distributed, thus variance can be assumed constant

```r
par(mfrow = c(2,2))
plot(model2)
```



Appendix E

```r
model <- lm(formula = mpg ~ am+ wt + qsec, data = mtcars)
summary(model)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## amManual     2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

```r
summary(model)$adj.r.squared
```

```
## [1] 0.8335561
```