Motor Trend Report

P.Y.Fong

12/31/2020

1.0 Executive Summary

In this report, the data set `mtcars` extracted from the 1974 *Motor Trend* US magazine, was used to explore the relationship between a set of predictor variables and the response variable, fuel consumption (MPG).
The data set comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). *Code Book Link*
Statistical Inference & Linear Regression methods were used 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"

The results of the analysis show that Manuals are significantly better for MPG. On the average Manuals are approx. 7.2 mpg better than Automatics in the data set.
However, fuel consumption (mpg) is negatively correlated to vehicle weight and positively correlated to 1/4 mile times regardless of transmission type. The model that quantifies the relationship between these key variables is:
MPG = 9.6 - 3.9(wt) + 1.2(qsec) + 3 (am: 0 for autos & 1 for manual)

However, residuals analysis show that further exploration is needed and should consider a non linear model with normalised data values and the exclusion of certain outlier data values.

2.0 Loading Libraries & Data.

```
library(ggplot2); library(GGally); data(mtcars); attach(mtcars)
```

3.0 Exploratory Data Analysis.

```
str(mtcars) # Data set structure
```

```
## '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 ...
```

Some variables were converted from numeric to factors.

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

3.1 Exploratory Observations.

1. (Refer A-1) Paired graph show high correlations between "mpg" with "disp", "hp", "wt" and "qsec". The predictors mentioned were also highly correlated among themselves.
2. The paired relationship between "mpg" and "am" warranted a closer look. (Refer A-2) Box plot, Manuals have higher values of MPG in general compared to Automatics.

4.0 Analysis with two sample T-test.

Null hypothesis: Mean MPG(auto) = mean MPG(manual), assuming normal distribution.

```
t.test(mpg ~ am)
```

```
##
##  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 0 mean in group 1
##        17.14737        24.39231
```

The null hypothesis was rejected. P-value < 0.05 (alpha = 5%). i.e. Mean MPG(auto) and mean MPG(manual) are significantly different. Also note that mean consumption (MPG) was about 7.2 mpg better for manuals.

5.0 Regression Analysis.

Using backward regression model fitting, a series of models were fitted with results (refer A-3) as follows:-

1. **fit 10. The all variable model.**
   Weak coefficient significance for all predictor variables i.e. p value > 0.05.
   F stat score 7.83. This model was not statistically significant.
   Adjusted R-squared = 0.0.779. The model explained approx 78% of the mpg variable.

2. **fit6.   The ignore factors model** factor variables "cyl", "vs", "gear" and "carb" removed.   Coefficient significance improved for "wt" and "qsec". P value > 0.05 remained for others.
   F stat score 27.09. This model was also not statistically significant.
   Adjusted R-squared = 0.8347. The model explained approx. 83% of the mpg variable.

3. **fit5. Reduced model** variable "drat" removed based on coefficient abs.t values < 1.0
   Coefficients of variables "wt", "qsec: and"am" become significant. P value < 0.05.
   F stat score 32.96. This model was still not statistically significant.
   Adjusted R-squared 0.8375. Slight improvement, explained approx. 84% of the mpg variable.

4. **fit3. Final model** variables "disp" and "hp" removed based on coefficient t values< 2.0
   Coefficients significant variables "wt", "qsec" and "am". P values < 0.05.
   F stat score 52.75. This model was somewhat statistically significant.
   Adjusted R-squared 0.8336 ie the model explained approx 84% of the mpg variable.

6.0 Model Scatter Plots. (refer A-4 & A-5)

The scatter plots for the final model fit3 show that "mpg" is negatively correlated to weight ("wt") and positively correlated to 1/4 mile times ("qsec") with manuals better at mpg in general.

7.0 Residuals Analysis. (refer A-6)

The Residuals vs Fitted plot showed a pattern suggesting that the linear model may not be ideal.

The Q-Q plot confirmed the assumption of a normally distributed data set.
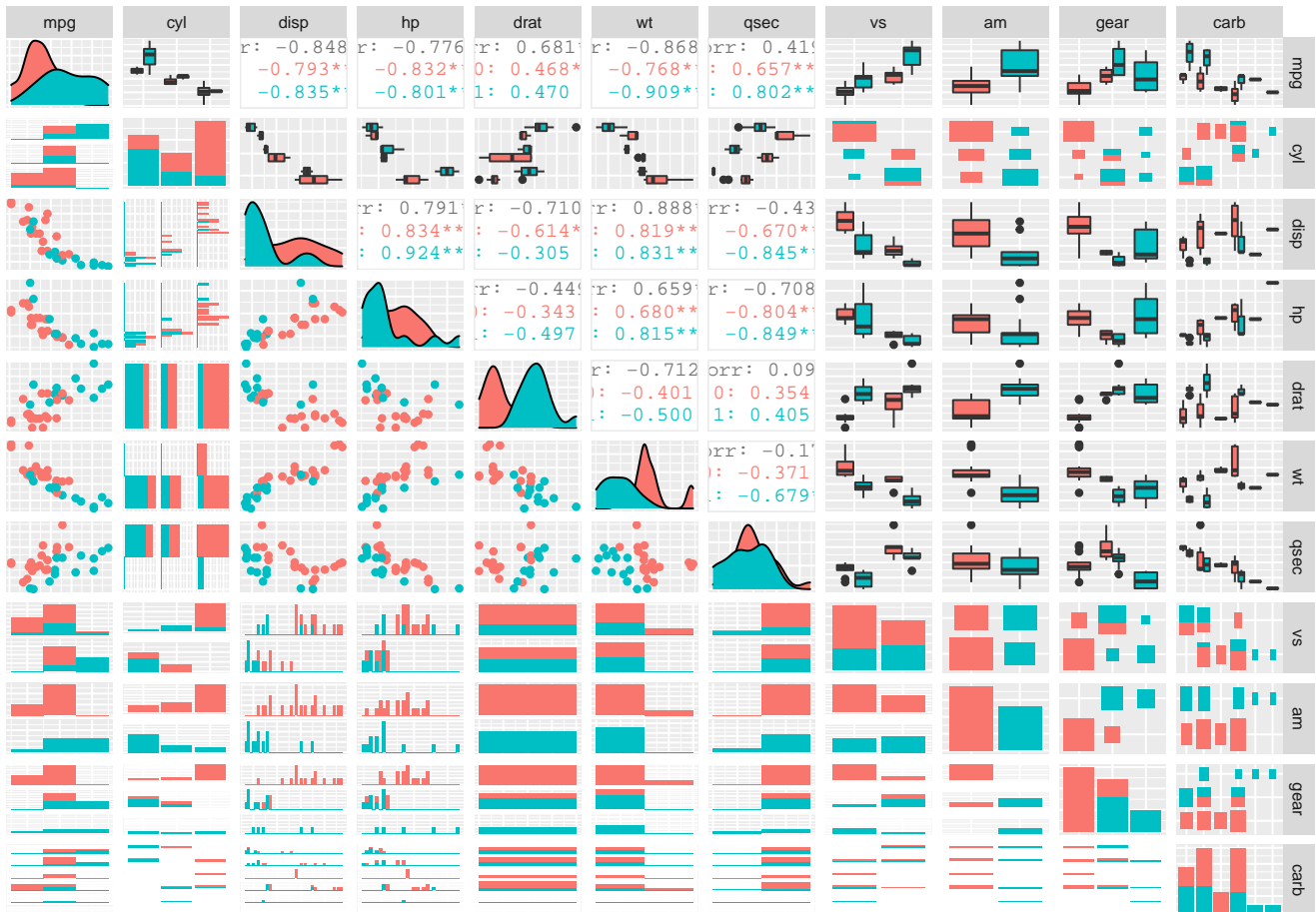
The Scale-Location plot showed a large spread with an angled fit line, so the assumption of equal variance was weak.

The Residuals vs Leverage showed that the model was influenced by a few outliers.

Appendices

A-1. Paired Graph

```
ggpairs(mtcars, aes(colour=am),axisLabels ="none", lower=list(combo=wrap("facethist", binwidth=10)))
```


A-2. Boxplot of MPG vs. Transmission.

```
ggplot(mtcars, aes(x=am, y=mpg, fill=am))+ geom_boxplot()+
  theme(axis.title.x= element_blank(),legend.position="none")+
  scale_x_discrete(labels = c('Manual','Automatic'))
```
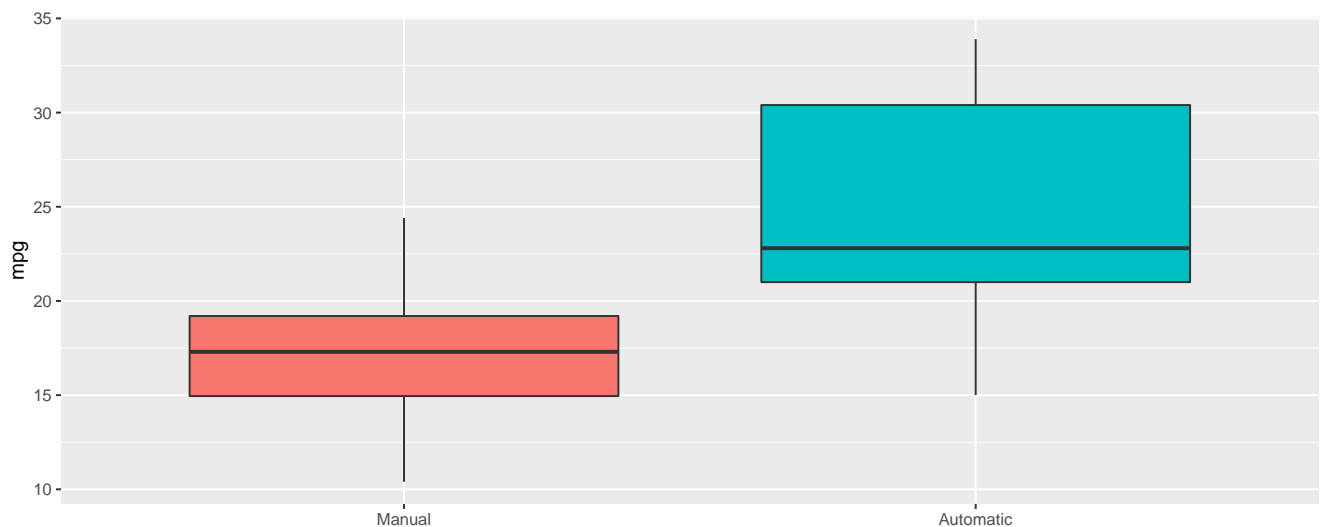
A-3. Regression Model Fitting

```r
fit10 <-  lm(mpg ~ ., data=mtcars); summary(fit10)          # all variables
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.87913   20.06582   1.190   0.2525
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## am1          1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb2       -0.97935    2.31797  -0.423   0.6787
## carb3        2.99964    4.29355   0.699   0.4955
## carb4        1.09142    4.44962   0.245   0.8096
## carb6        4.47757    6.38406   0.701   0.4938
## carb8        7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

```r
fit6 <-  update(fit10,. ~ .-cyl-vs-gear-carb); summary(fit6)  # less cyl, vs, gear & carb
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + drat + wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2669 -1.6148 -0.2585  1.1220  4.5564
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.71062   10.97539   0.976  0.33848
## disp         0.01310    0.01098   1.193  0.24405
## hp          -0.02180    0.01465  -1.488  0.14938
## drat         1.02065    1.36748   0.746  0.46240
## wt          -4.04454    1.20558  -3.355  0.00254 **
## qsec         0.99073    0.48002   2.064  0.04955 *
## am1          2.98469    1.63382   1.827  0.07969 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.45 on 25 degrees of freedom
```

```
## Multiple R-squared:  0.8667, Adjusted R-squared:  0.8347
## F-statistic: 27.09 on 6 and 25 DF,  p-value: 8.637e-10
```

```r
fit5 <-  update(fit6,. ~ .-drat); summary(fit5)               # less drat
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5399 -1.7398 -0.3196  1.1676  4.5534
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.36190    9.74079   1.474  0.15238
## disp         0.01124    0.01060   1.060  0.29897
## hp          -0.02117    0.01450  -1.460  0.15639
## wt          -4.08433    1.19410  -3.420  0.00208 **
## qsec         1.00690    0.47543   2.118  0.04391 *
## am1          3.47045    1.48578   2.336  0.02749 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.429 on 26 degrees of freedom
## Multiple R-squared:  0.8637, Adjusted R-squared:  0.8375
## F-statistic: 32.96 on 5 and 26 DF,  p-value: 1.844e-10
```

```r
fit3 <-  update(fit5,. ~ .-disp-hp); summary(fit3)            # less disp & hp
```

```
##
## 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 ***
## am1           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
```

A-4. Scatter Plot of MPG vs. Weight by Transmission

```
ggplot(mtcars, aes(x=wt, y=mpg, group=am, color=am))+ geom_point()+
  stat_smooth(method = "lm")+ xlab("Weight")+
  scale_colour_discrete(labels=c("Automatic", "Manual"))
```
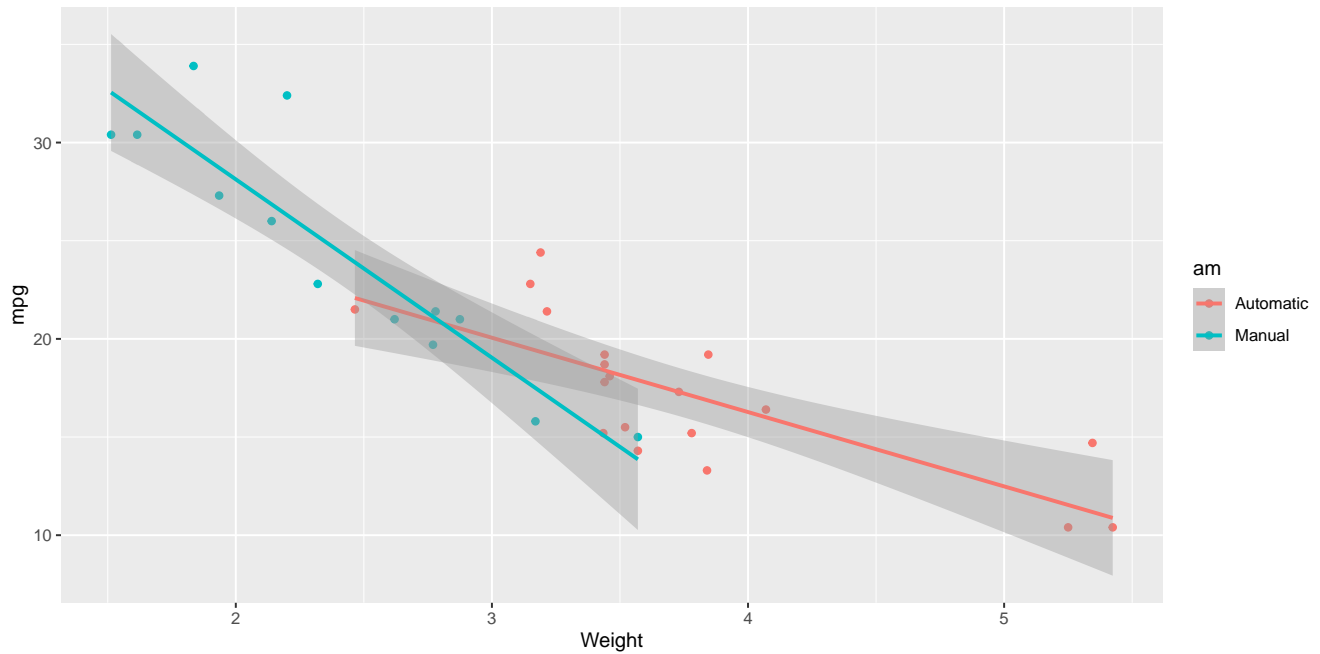

A-5. Scatter Plot of MPG vs. Quarter Mile Time

```
ggplot(mtcars, aes(x=qsec, y=mpg, group=am, color=am))+ geom_point()+
  stat_smooth(method = "lm")+ xlab("Weight")+
  scale_colour_discrete(labels=c("Automatic", "Manual"))
```
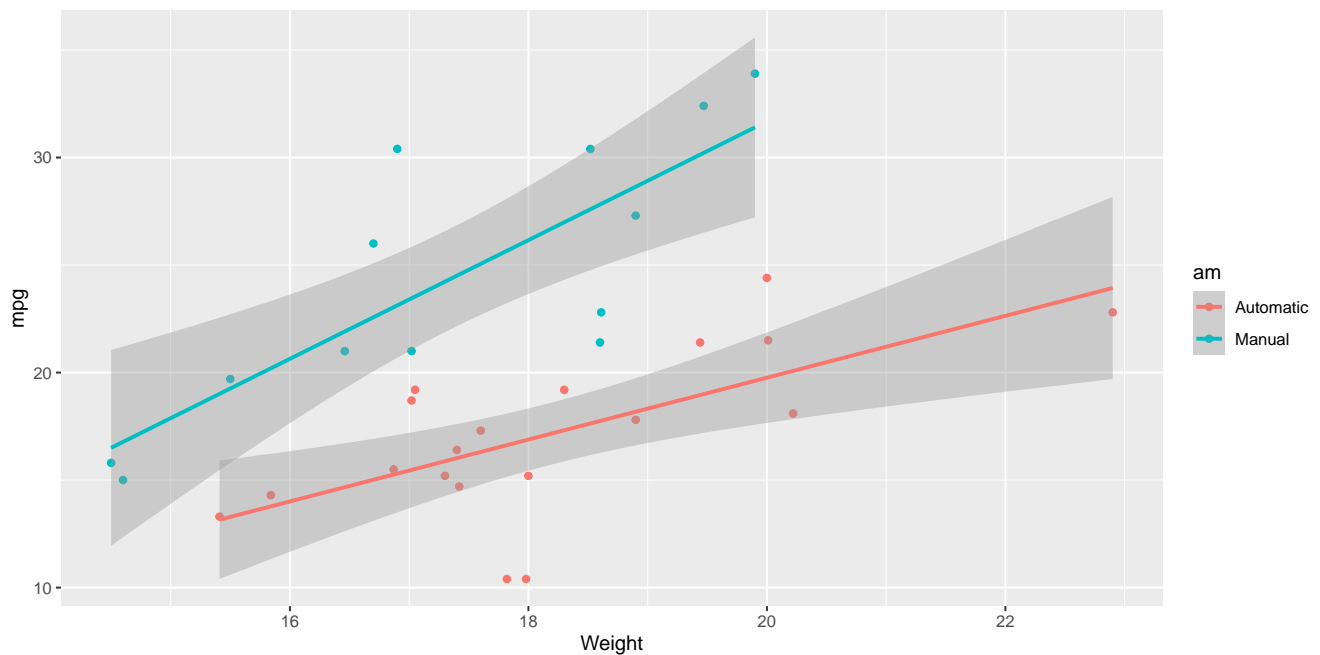
A-6. Residual Plots

```
library (ggfortify); autoplot(fit3)
```