Motor Trend Data Evaluation

Regression Model Project Assignment

Executive Summary

In this project, we analyze the data from Motor Trend to answer the following qustions and provide additional observations as the data analysis warranted.

1. Is an automatic or manual transmission better for mpg?
    - Based on the data on hand, manual transmission does perform better than automatic transmision given all other factors are equal.
2. Quantifying how different is the mpg between 'Automatic' and 'Manual' transmissions?
    - In average, the manual transmission is at approximately 7 miles per gallon advantages compare to autmatic transmision. Howerever, other factors of automobile, such as weight, cylinder, and horse power, can confound the miliage per gallon performance. Based on the modeling result below, those factors are significantly influnce the mpg results.

Data Loading and Transforming

As review the data, there are 32 observations with 11 variables. All the data are in numberic format, we are going to transform those variables known to be the characteristics of an autombile by factoing the data.

```
## '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 : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ 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  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

Exploratory Data Analysis

Since the primary interest is about the transmission performance evaluated by miles per gallon. We first plot the boxplot in Appendix 1 to explore the difference between automatic and manual. As the plot shows the amount of differences is significant. We decide to run a t-test between auto and manual transmission against mpg assume that the variance are the same.

```
##
##  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
```

The p-value in the t-test is 0.0014 signifincant less than 0.05, we reject the null hypothesis that there is not differences between types of transmission. However, there are factors in the data set that is interesting to us, we plot a pairwise scatterplot to further ecplore the relationship between 'mpg' and other variables.

As seen in Appendix 2, the cyl, disp, hp and wt seems have either positive or negative impact to the mpg. We will explore those variable relationship against mpg in the following analysis while we develop the linear models.

Regression Analysis

In this section, we will build linear regression models by applying stepwise method to identify the best model

Establish the models Include, the base model(model_0), initial model(model_1), and best model(model_c).

```
model_0 <- lm(mpg ~ am, data = mtcars)
model_1 <- lm(mpg ~ ., data = mtcars)
model_c <- step(model_1, direction = "backward")
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

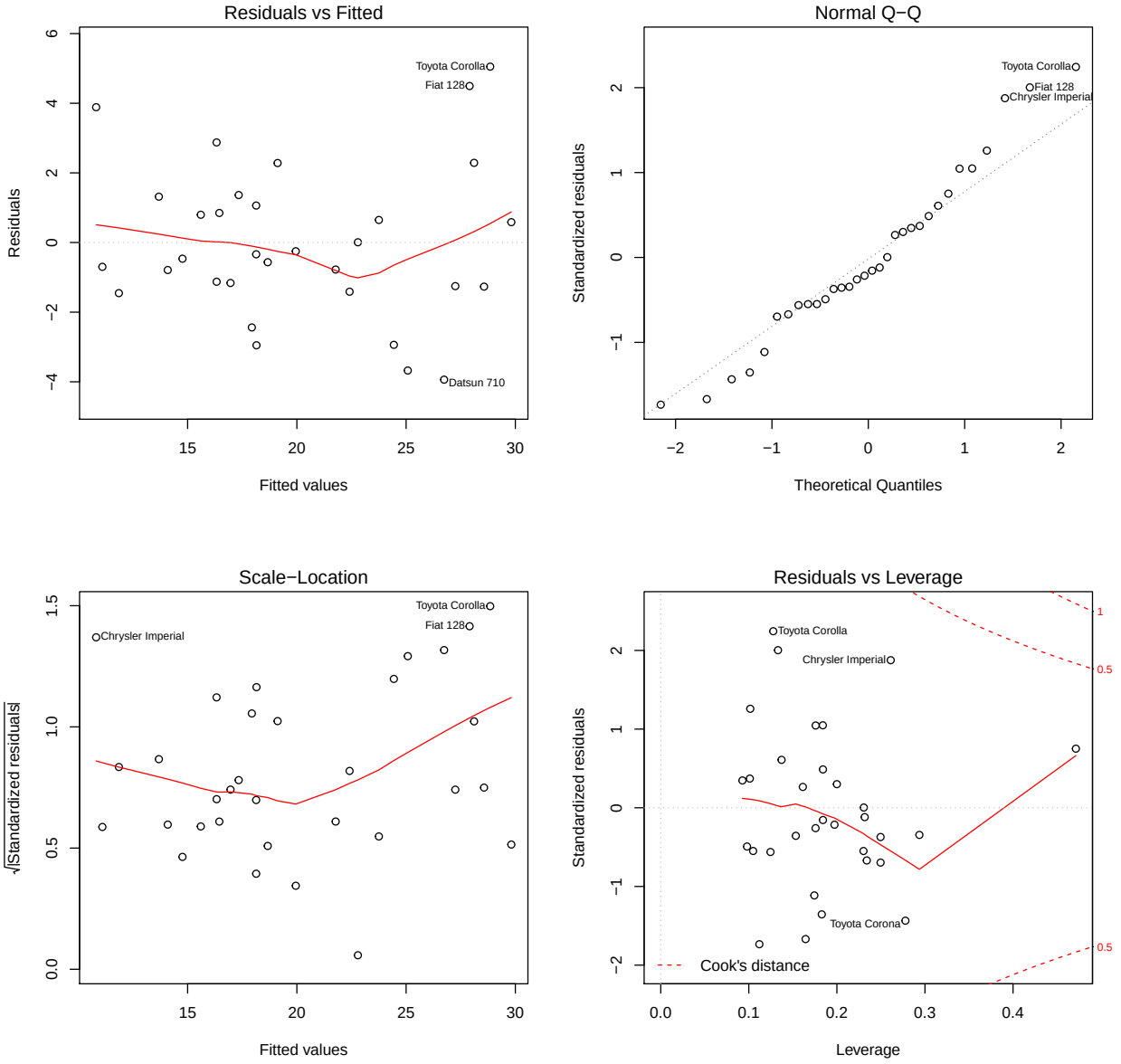
As show in the results above, with R-square = 0.84, the selected variables have interprete the 84% variablity among all variables. In next step, we will evaluate the base model agianst choice model to see if the choice model is better performer than the base model.

```r
anova(model_0, model_c)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residuals Analysi and Diagnostics Residual Analysis is performed of the model of choice to see the fitness of the suggested model, also to identify the potential outliers to readjust the model.

Figure 1. residuals analysis plots



Conclusions:

- The Residuals vs Fitted plot verify the randomness of the data collected.
- The residual data are nomorlly distributed.
- In the Scale-Location plot the points are in a constant band pattern, indicating constant variance.
- The Residuals vs Leverage plot shows the concern of Toyota Corolla, Chrysler Imperial, and Toyota Corona.

Here computes the regression diagnostics of the choice model to find out the results of leverage points.

```
lev <- hatvalues(model_c)
tail(sort(lev),3)
```

```
##       Toyota Corona Lincoln Continental       Maserati Bora
```

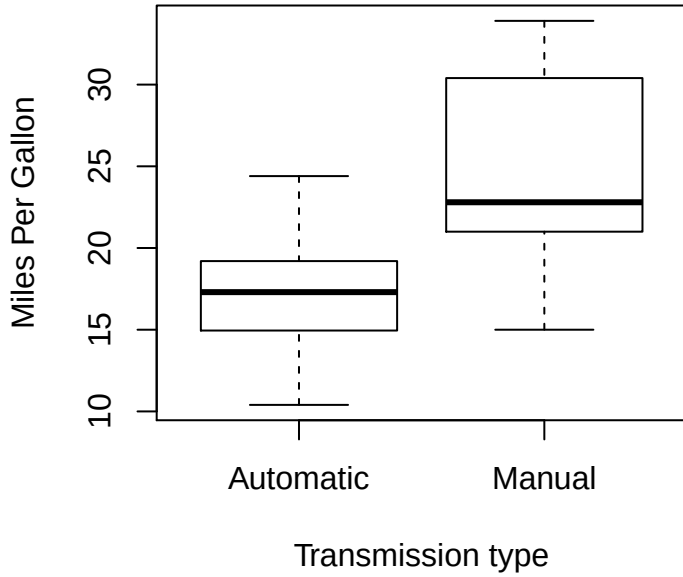
```
##           0.2777872           0.2936819           0.4713671
```

```
inf <- dfbetas(model_c)
tail(sort(inf[,6]),3)
```

```
## Chrysler Imperial          Fiat 128     Toyota Corona
##         0.3507458         0.4292043         0.7305402
```

Conclusions

- Miles per gallon(mpg) potentially will increase by 1.81 more with 'Manual' transmission than 'Automatic' transmission
- Miles per gallon (mpg) potentially will decrease by 2.5 for every 1000 lb of increase.
- Miles per gallon (mpg) decreases with increase of hp.
- Miles per gallon(mpg) will decrease by a factor of 3 and 2.2 if number of cylinders cyl increases from 4 to 6 and 8, respectively (adjusted by hp, wt, and am).



Appendix Appendix 1. Boxplot for mpg by type of transmission

Appeneix 2. Scatterplot for all Variables in mtcars

Motor Trend Cars Scatterplot