Regression Models Course Project

Author: **Fred Zhou**

In this document, we will try to answer the following questions:

- Q1: "Is an automatic or manual transmission better for MPG"
- Q2: "Quantify the MPG difference between automatic and manual transmissions"

By default, we assume that for the `mpg`, the lower the value the better.

(For `am`, `0` for automatic transmission, `1` for manual transmission.)

Q1. Is an automatic or manual transmission better for MPG

To answer this question, we assume that the all the variables in the population follow normal distribution. Thus we first use Student's T test to address whehter there's difference in these two groups

```
test_mpg=t.test(mtcars$mpg[mtcars$am==1],mtcars$mpg[mtcars$am==0])
print(test_mpg)
```

Student's T-test between AUTOMATIC and MANUAL (alpha=0.05)

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg[mtcars$am == 1] and mtcars$mpg[mtcars$am == 0]
## 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:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

```
print(paste(
'The P-value for the T-test between AUTOMATIC and MANUAL transmissions for the mpg is ',
round(test_mpg$p.value,digits = 4),sep=''))
```

```
## [1] "The P-value for the T-test between AUTOMATIC and MANUAL transmissions for the mpg is 0.0014"
```

```
print(paste(
'Mean value for the mpg with AUTOMATIC transmissions:',
round(test_mpg$estimate[1],digits = 2),sep=''))
```

```
## [1] "Mean value for the mpg with AUTOMATIC transmissions:24.39"
```

```
print(paste(
'Mean value for the mpg with MANUAL transmissions:'
,round(test_mpg$estimate[2],digits = 2),sep=''))
```

```
## [1] "Mean value for the mpg with MANUAL transmissions:17.15"
```

Thus we could address that indeed the types of transmission will affect the `mpg`, and on average `AUTOMATIC` will bear a *higher consumption of fuel* against the `MANUAL` transmission, and the average difference is around *7.24* miles per Gallon used.

Q2. Quantify the MPG difference between automatic and manual transmissions

```
sort(abs(cor(mtcars)[1,]))
```

Correlation analysis winthin all variables against the mpg

```
##      qsec      gear      carb        am        vs      drat        hp
## 0.4186840 0.4802848 0.5509251 0.5998324 0.6640389 0.6811719 0.7761684
##      disp       cyl        wt       mpg
## 0.8475514 0.8521620 0.8676594 1.0000000
```

We already get the hint that the `AUTOMATIC/MANUAL` have impacts on the fuel consumption, thus from the correlation analsis we could guess that any variant with a higher correlation value against `AUTOMATIC/MANUAL` may contribute to the fuel consumption. including:

1.`vs` - V/S

2.`drat` - Rear axle ratio

3.`hp` - Gross horsepower

4.`disp` - Displacement (cu.in.)

5.`cyl` - Number of cylinders

6.`wt` - Weight (1000 lbs)

Thus, we could guess that it's reasonable to include any variable into the linear regressions. We could make a most general form of regression, then add in more variants to further optimize our model.

General model We only take the `am` as variables to do the linear regression first:

```
fit_1 <- lm(mpg~am, data = mtcars)
summary(fit_1)
```

```
##
## 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 ***
## am             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
```

Based on the stat data we could address:

- On average, AUTOMATIC car have 17.15 MPG and MANUAL transmission cars have 7.25 MPG more
- The R^2 value is only 0.36, which means that our current model only explains 36% of the variance

```r
fit_2 = step(lm(data = mtcars, mpg ~ .),trace=0,steps=50000)
summary(fit_2)
```

Multivariate model - adapted selection of variants

```
##
## 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 ***
## am            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
```

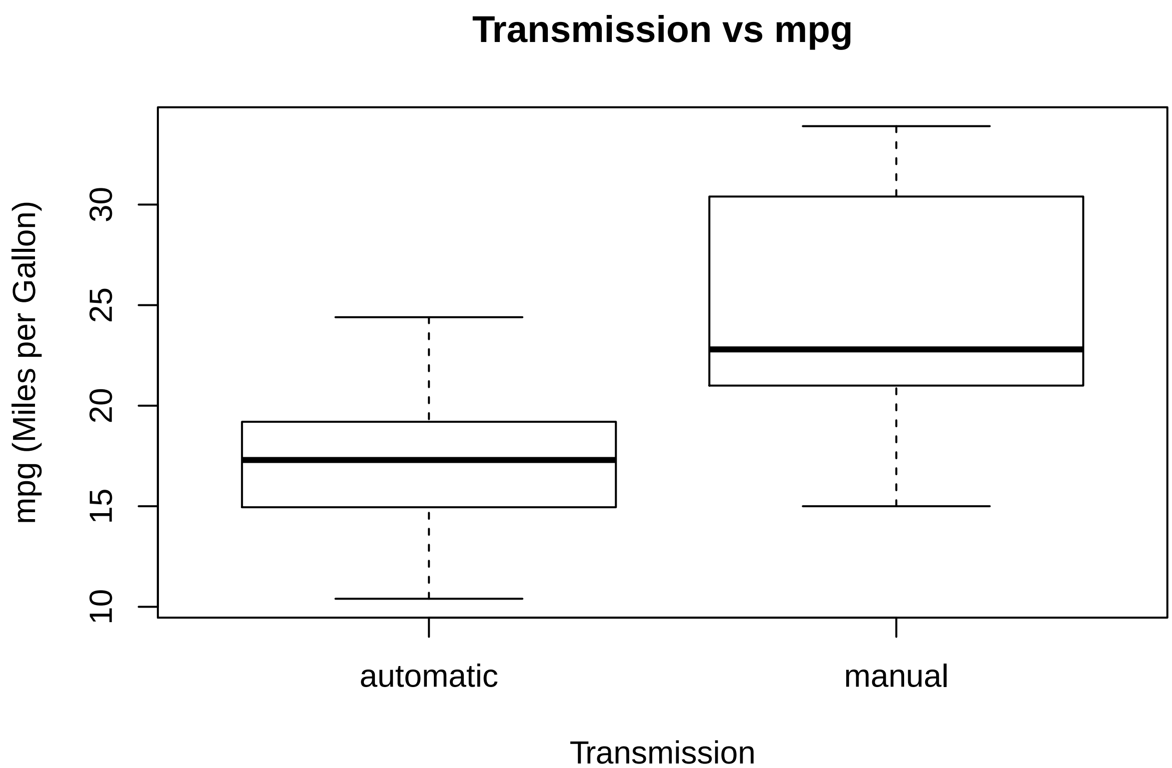
Summary

This model explains 84% of the variance in miles per gallon (`mpg`), which is accaptable for the predicion of `mpg` with new data. Based on the multivariate model we could address:

- `MANUAL` is beneficial for the fuel saving, after model adjusting the value comes to be *2.936* miles per gallon.
- `wt` affect huge against the `mpg`, which is appearant since more load will eventually consume more fuel.

APPENDIX

Visualize the data between AUTOMATIC and MANUAL



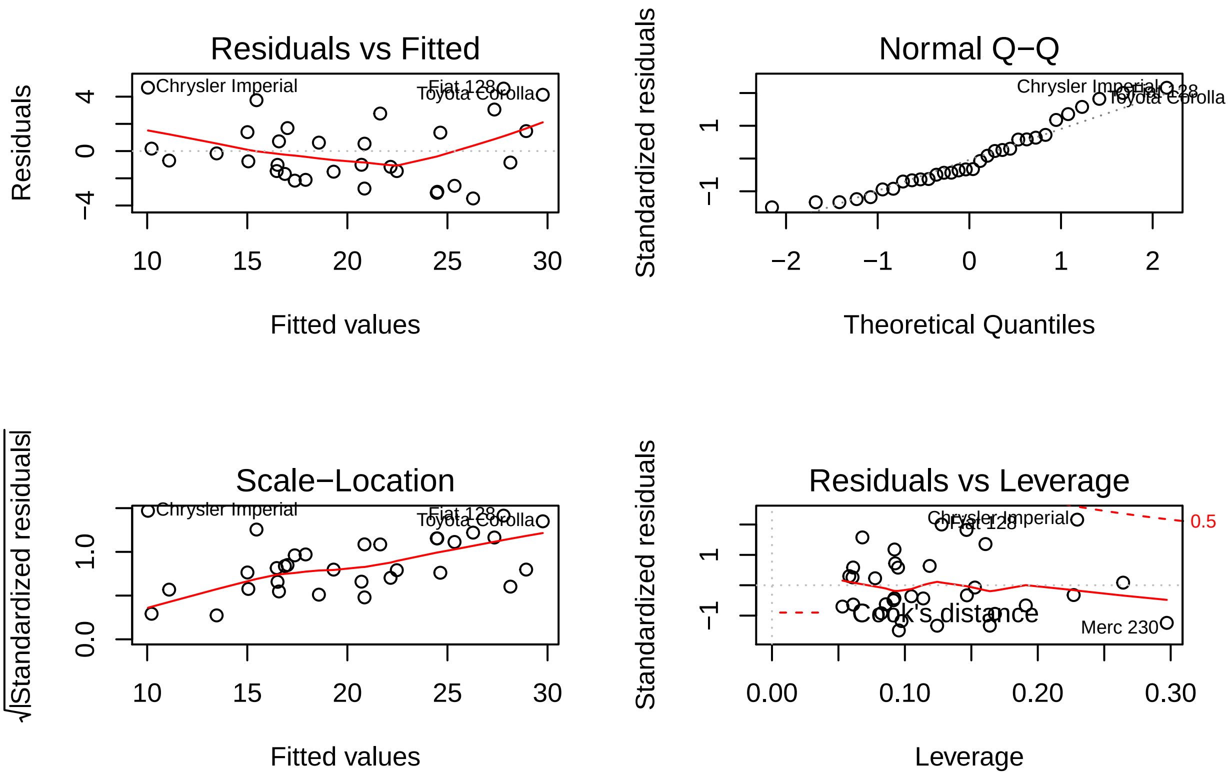
Comparision of general and multivariate model 1.ANOVA

```
anova(fit_2, fit_1)
```

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

2.Residual diagnostics for multivariate model

```
par(mfrow = c(2,2))
plot(fit_2)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
Merc 230
0.5