Regression Models Course Project

Executive Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

Data Loading and processing

I load the dataset and run some summary statistics on dataset.

```
library(datasets)
data(mtcars)
head(mtcars)
```

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

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

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

I transform the "am" variable to factro variable with two levels of "Automatic" and "Manual".

```
mtcars$am<-factor(mtcars$am, labels=c('Automatic', 'Manual'))
```

The boxplot of MPG by transmission types (Appendix 2) shows that manual cars are better im miles per gallon than the automatic cars. Manual cars use one gallon of gas per 22.8 miles, while automatic cars use one gallon of gas for every 17.3 miles.

```
median(mtcars$mpg[mtcars$am=="Manual"])
```

```
## [1] 22.8
```

```
median(mtcars$mpg[mtcars$am=="Automatic"])
```

```
## [1] 17.3
```

The "t-test"" tests the hypothesis that if the true difference in means of automatic and manual cars are equal to zero or not.

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

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$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 of 0.001374 is less than 0.05 and I reject the null hypothesis that the true difference in means is equal to 0 (mean of atomatic cars is different than the manual cars). Based on the sample estimates of mean in cars, we can conclude that mannual cars are better in MPG than the automatic cars, with all other conditions constant.

2. Quantify the MPG difference between automatic and manual transmissions

In this part, I quantify the MPG difference between automatic and manual transmissions, and find out if other variables also explain the MPG differences or not.

```
fit<-lm(mpg~., data=mtcars)
summary(fit)$coef
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## amManual     2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

The only significant coefficient in multivariate linear regression is weight. Including all varaibles in the model overfit the results. Then, I use the step function (automatic variable selection). The following model chooses the best linear regression models.

```
bestfit<-step(lm(mpg~., data=mtcars))
```

```
summary(bestfit)
```

```
##
## 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
```

The results show that the best model includes wt, qsec, and amManual variables. All the coefficients except intercept are statistically significant and different than 0 (overall p-value<0.01). The model explains 84.97% of the variance.
For every lb/1000 increase in weight of car, mpg decreases about 4. For every quarter mile time the mpg increases by 1.2. The manual cars are 2.9 mpg better than automatic cars. The residual plots scattered randomly (Appendix 2).

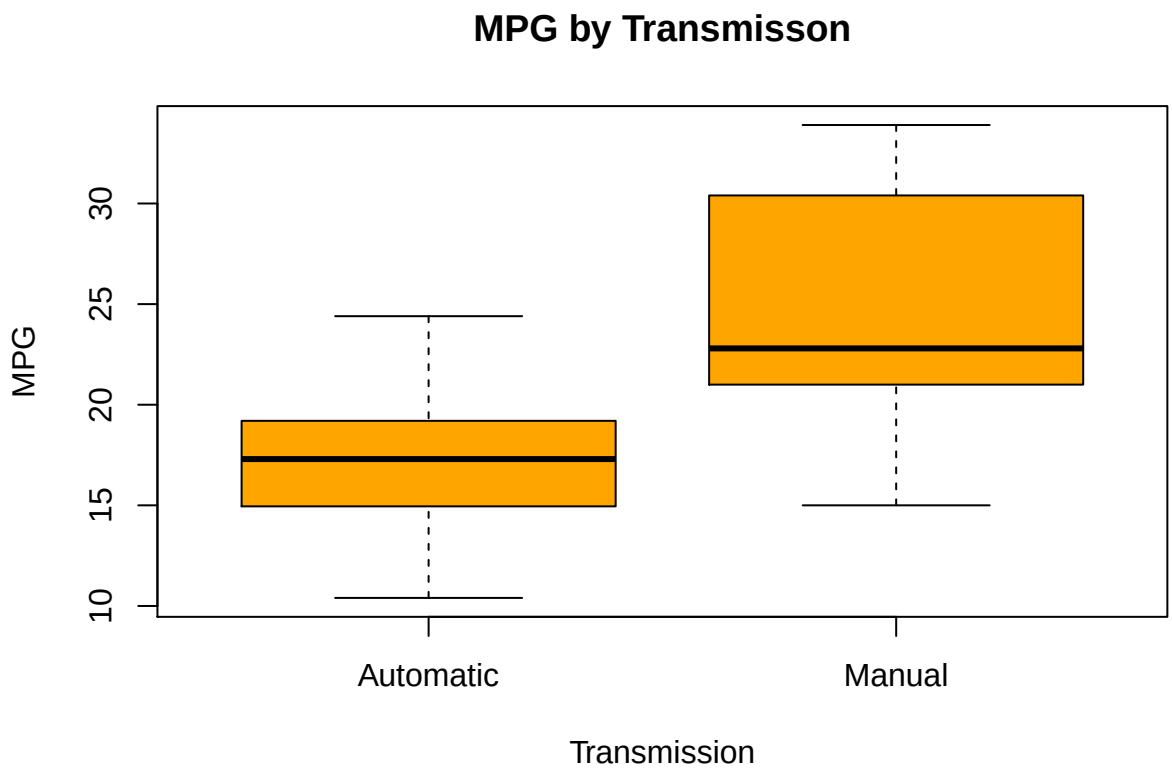
Conclusion

Manual transmission cars are better in mpg than the automatic cars by 2.93 mpg. The other important factors beside the transmission types are weight, and quarter mile time.

Appendix

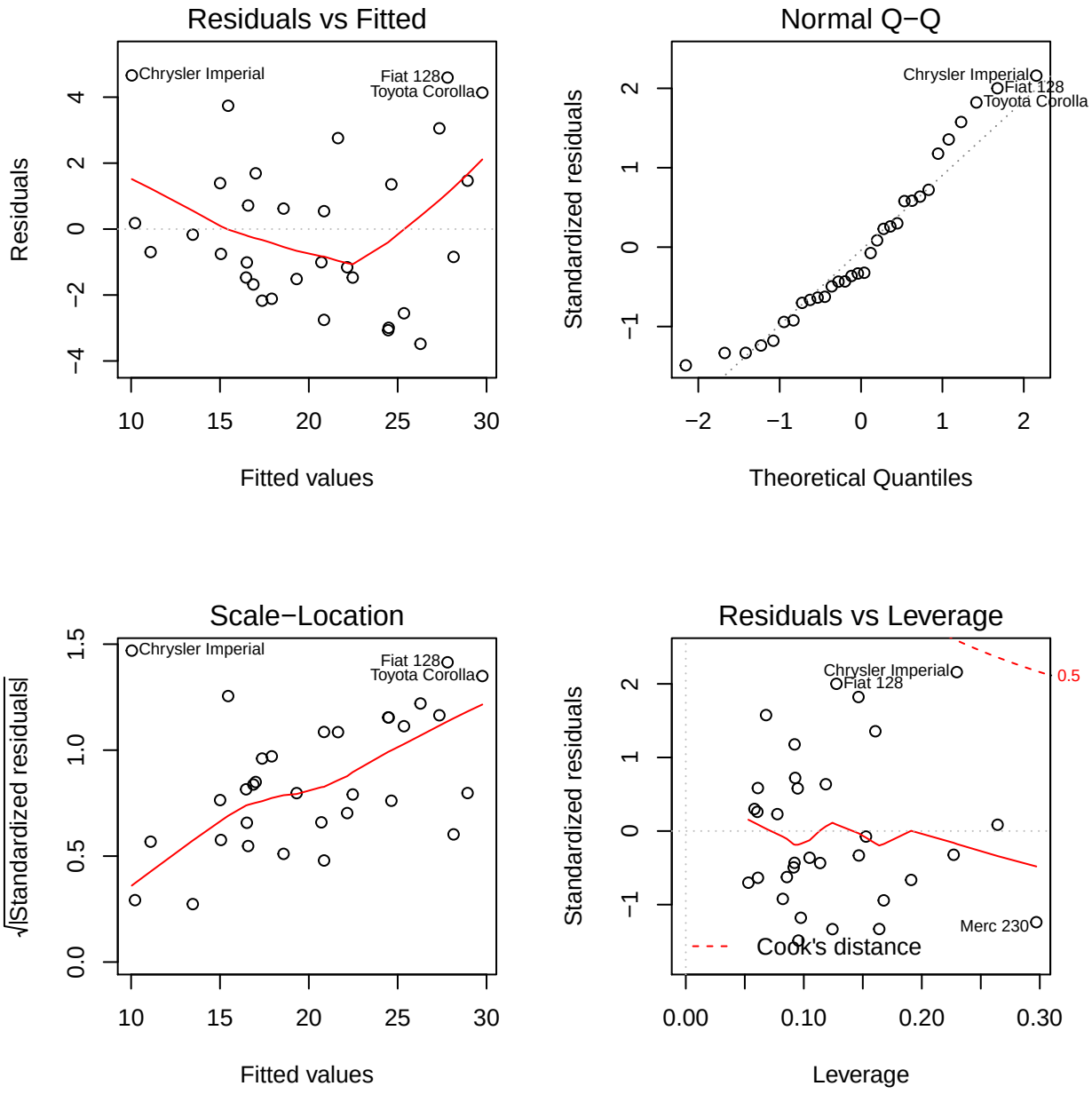
Appendix 1

```
boxplot(mtcars$mpg ~ mtcars$am, data=mtcars, col="orange",
        xlab="Transmission",
        ylab="MPG",
        main="MPG by Transmisson")
```



Appendix 2

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


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