Regression Models Project Report

Haonan

October 28, 2016

Excutive Summary

This report works for “Motor Trend” Magizine about automobile industry. The main goal of this report is particularlly to figure out the relationship between MPG and its influential factors. Specifically, the differenct types of transmission(auto and manual). Anfter the data analysis process, this report concluded that at the excutive level, switching automatic transmission to manual transmission, purchase a lighter vehicle, and don’t desire for high horsepower will definitely help the car owner to make their MPG looks better.

First Glance of the Data

The dataset used in this report is called “mtcars”, sourced from Henderson and Velleman (1981), and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). For more details of the dataset, please kindly type “?mtcars” in the R console. Some of the aspects are factors, thus this report will firstly transfer these aspects to characterized factors using R codes for a more convenient further analysis.

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
attach(mtcars)
```

Basic Analysis

The first step to conduct a basic exploratory data analysis is to plot the relationship between MPG and types of transmission. As we could see from graph1 (See Appendix Graph1), there seems a obvious difference between manual and auto transmission. We use t test to verify this hypothesis:

```
tt <- t.test(mpg ~ am)
tt$p.value
tt$estimate[2] - tt$estimate[1]
```

Statistically, we should concern the p value and the mean of difference. As the result shows above, p value is less than 1% indicates a significant difference thus we should reject the null hypothesis which assumes they are equal. and the mean difference is 7.24.

Clearly MPG will not solely rely on only transmission. We also plotted a 3*3 graph showing the scatterplot of MPG with all other aspects (See Appendix Graph2). The graphs seems like all the aspects have some relationship with MPG. Aftering absorbing these infomation we’ll pursue to the regression analysis stage.

Regression Analysis

Fitting Model with All Aspects

We fitted a regression model with every aspect and all of its two-way interactions:

```
fitall2 <- lm(mpg ~ .^2, mtcars)
sumfitall2 <- summary(fitall2)
sumfitall2$sigma
sumfitall2$r.squared
```

Residual std error is missing, but the R-squared value is 1 which indicates a perfect fitting model. Take a look at the coefficients, the estimate value is subtle and p values are missing. Containing too many unnecessary variables does not make a good fitting model, we'll eliminate some later on.

Fitting Model with the Best Aspects

Since considering two-way interactions will make this model too complicated and hence make it uninterpretable and meaningless, this report will only consider necessary interactions. firstly we'll fit the non-interaction model namely "fitall", then use Stepwise Algorithm (step() function) to choose the best aspects in the linear model, and next consider some interactions among the chosen variables.

```
fitall <- lm(mpg ~ ., mtcars)
fitstep <- step(fitall, direction = "both")
```

We've find out that hp and wt are significant(See Appendix Table3), then we could use these variables to construct the best fitting model:

```
fitbest <- lm(mpg ~ (am + wt + hp)^2)
summary(fitbest)$sigma
summary(fitbest)$df
summary(fitbest)$r.squared
anova(fitstep, fitbest)
```

(See Appendix Table4 for the detailed summary) Residual std error is 2.18466 under 25 degree of freedom. R square value is 0.8686 indicates a good fitting model. For the coefficients, the model provides am1 as 4.26, am1:wt as -2.77 and am1:hp as 0.027. Here's the exolaination: when changing from automatic transmission to manual transmission while other variables keep constant, the MPG will be 4.26 - 2.77*wt* + *0.027*hp higher. Definitely indicates that the MPG will be better when using manual transmission, but the heavier the vehicle, the lower efficacy of changing transmission.

Through the ANOVA test, the p-value of the second model is less than 5%, which indicate the best fitting model (fitbest) is significant, we'll accept this model.

Residual Diagnostics

(Please see Appendix Graph 5)Based on the first plot of Graph5, we have the residuals vs fitted plot. This is a messy scatterplot, indicating that there's seemed no pattern of the residuals. But we could indified there might exist 2 outliers.We could test its leverage using hatvalue function:

```
tail(sort(hatvalues(fitbest)),2)
```

```
##        28        31
## 0.6234270 0.7226269
```

In Conclusion

1: There does have a difference automatic transmision and manual transmission, the difference in the mean of MPG is 7.245

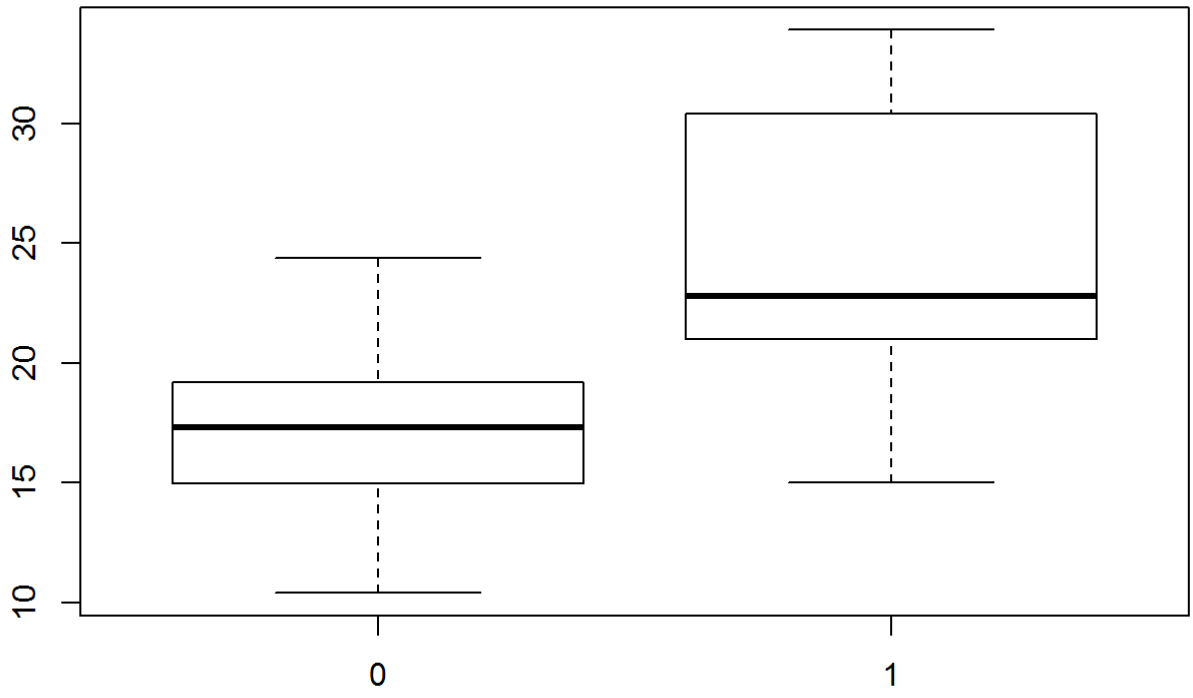
2: Based on the regression model, MPG = 46.69 + 4.26(am1) - 6.74(wt) -0.12(hp) - 2.77(am1*wt) + 0.03(am1*hp) + 0.02(wt*hp). Switching from auto transmission to manual transmission will increas MPG by 4.26 - 2.77*wt + 0.03*hp.

3: Every 1000 lbs increase in weight will decrease MPG by 6.74 + 2.77*am1 + 0.02*hp.

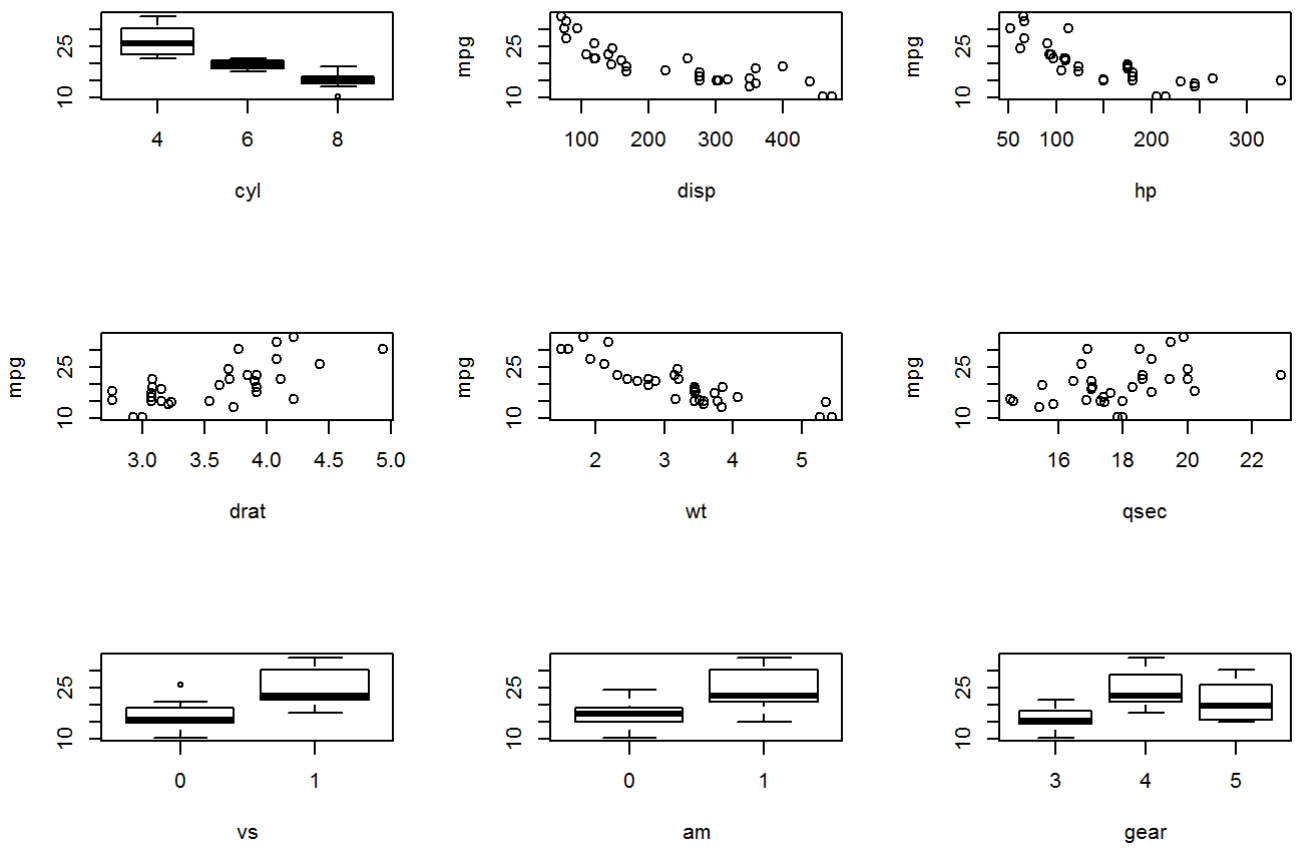
4: Every 1 hoursepower grows, the MPG will decrease by 0.12 - 0.03*am1 -0.02*wt

5: The heavier the vehicle, the lower efficacy of switching from auto transmission to manual transmission.

Appendix



Graph1: Boxplot of MPG and types of transmission



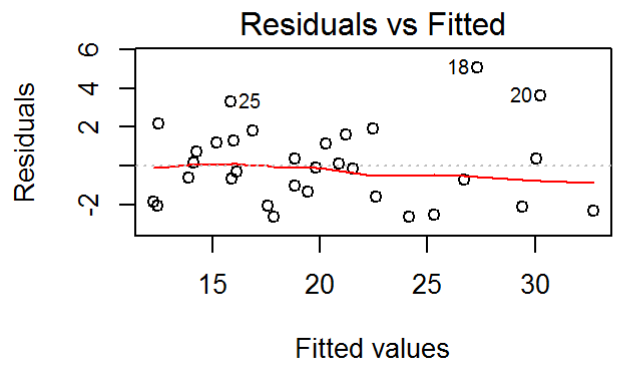
Graph2: Scatterplot of MPG and all other aspects

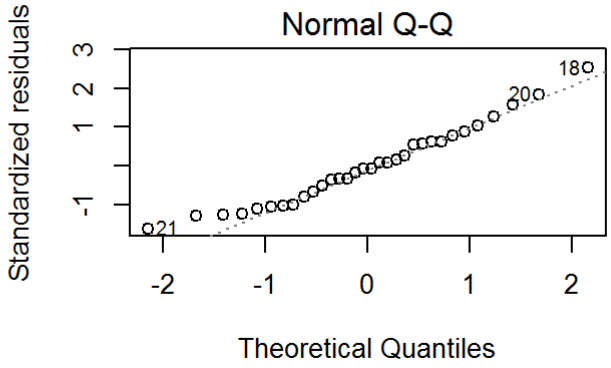
```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## am1          1.80921138 1.39630450  1.295714 2.064597e-01
```

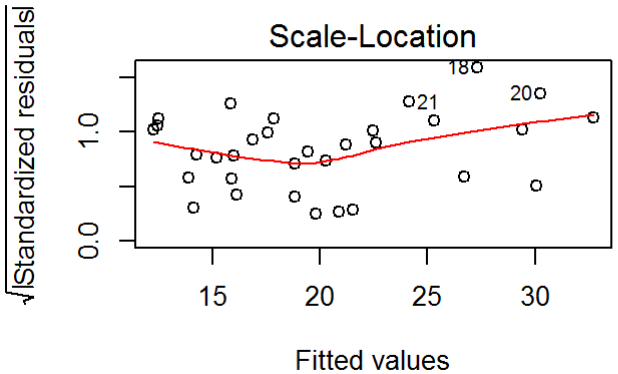
Table3: Coefficients for “fitstep” model

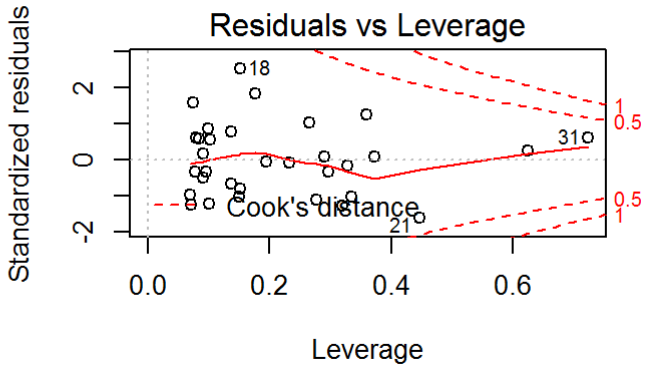
```
##
## Call:
## lm(formula = mpg ~ (am + wt + hp)^2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6537 -1.6815 -0.1368  1.2373  5.0885
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46.68556    8.94711   5.218 2.12e-05 ***
## am1          4.25637    6.49419   0.655   0.5182
## wt          -6.74000    2.77159  -2.432   0.0225 *
## hp          -0.12034    0.04454  -2.702   0.0122 *
## am1:wt      -2.77304    2.55022  -1.087   0.2872
## am1:hp       0.02658    0.01837   1.447   0.1602
## wt:hp        0.02401    0.01288   1.864   0.0741 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.185 on 25 degrees of freedom
## Multiple R-squared:  0.894,  Adjusted R-squared:  0.8686
## F-statistic: 35.16 on 6 and 25 DF,  p-value: 5.193e-11
```

Table4: Summary of “fitbest” model









Graph5: Plot of the best fitting model, especially Residual Plot