Motor Trend Analysis

Hsin Yu Cheng May 24, 2015

Executive Summary

Motor Trend, a magazine about the automobile industry, is interested in the relationship between a set of variables and miles per gallon. The purpose of the data analysis is to see whether miles per gallon are significantly different when people use automatic and manual transmission. For the final result, statistics supoorts that manual transmission is better than automatic transmission in miles per gallon.

Load Data and check data information

```
data(mtcars)
str(mtcars)
```

"mtcars" includes 32 observations and 11 variables.

```
mtcars <- mtcars %>% mutate(Transmission = ifelse(am == 1, "manual", "automatic"))
```

Exploratory analysis

Scatterplot between variables.

```
pairs(mtcars)
```

1.See Figure 1 in appendix.
2.It shows relationships between variables.

Is an automatic or manual transmission better for MPG?

```
ggplot(mtcars, aes(x = Transmission, y = mpg, colour = Transmission)) +
  geom_boxplot() + theme_bw() + ylab('Miles/(US) gallon') +
  theme(legend.position = "top") +
  ggtitle("Miles/(US) gallon between automatic and manual") +
  theme(plot.title = element_text(lineheight = .8, face = "bold", size = 12))
```

1.See Figure 2 in appendix.
2.Visualization shows that manual transmission is better than automatic transmission.

Quantify the MPG difference between automatic and manual transmissions?

```
mean <- mtcars %>% group_by(Transmission) %>% summarise(Transmission_Mean = mean(mpg))
```

Statistics Evidence

```
t.test(mpg ~ Transmission, mtcars)
##
##  Welch Two Sample t-test
##
## data:  mpg by Transmission
## 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
```

Interpretation :
1.From the result of t-test, the p value is 0.001374 < 0.05. It achieves significant level. It means that there is significant difference among manual and automatic transmission on dependent variable "Miles per gallon".
2.Miles per gallon of **manual transmission** is better than that of **automatic transmission**, 24.39 and 17.15 respectively.

Regression Model

1. Variable selection : Use "backward method"

```
mydata <- mtcars %>% mutate(cyl = as.factor(cyl), vs = as.factor(vs), gear =
as.factor(gear), carb = as.factor(carb)) %>% select(-am)

full.model <- lm(mpg ~ ., mydata)
reduced.model <- step(full.model, direction = "backward")
```

It shows that the best model includes four variables, namely Transmission(am), hp, cyl and wt.

2. Best model

```
Best.model <- lm(mpg~ Transmission + hp + wt + cyl, data = mydata)
summary(Best.model)
## Call:
## lm(formula = mpg ~ Transmission + hp + wt + cyl, data = mydata)
##
## 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 ***
## Transmissionmanual  1.80921    1.39630   1.296  0.20646
## hp                 -0.03211    0.01369  -2.345  0.02693 *
## wt                 -2.49683    0.88559  -2.819  0.00908 **
## cyl6               -3.03134    1.40728  -2.154  0.04068 *
## cyl8               -2.16368    2.28425  -0.947  0.35225
## ---
## 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
AIC(full.model)
## [1] 169.2155
AIC(Best.model)
## [1] 154.4669
```

Interpretation

1.The value of **Adjusted R-squared is 84.01%**. It represents that 84.01% of the total variance in the "mpg(Miles per gallon)" around its mean is accounted for by Transmission, hp, wt and cyl variables.

2.For hp, wt and cy16 variables, the p values are 0.02693 < 0.05, 0.00908 < 0.05, and 0.04068<0.05. They all **achieve significant level** and show negative relationship with mpg.

3.For hp(horsepower), the coefficient -0.03211 indicates that for every additional **horsepower** you can expect **mpg to decrease by an average of 0.03211 miles per gallon.**

4.For wt(Weight), the coefficient -2.49683 indicates that for every additional **Weight(lb/1000)** you can expect **mpg to decrease by an average of 2.49683 miles per gallon.**

5.For cyl6(6 cylinders), compared to 4 cylinders(the reference level), the coefficient -3.03134 indicates that 6 cylinders is less than 4 cylinders in miles per gallon at 3.03134.

6.**AIC of best model improves from 154.47(full model) to 169.22 and adjusted R-squared of best model improves from 77.9%(full model) to 84.01%.**

Regression Residual Plots of Best Model.

```
par(mfrow=c(2, 2))
plot(Best.model)
```

1.See Figure 3 in appendix.

2.**Interpretation : The residuals are normally distributed and homoscedastic.**

Diagnostics : Exam influential points.

```
library(car)
influence.measures(Best.model)
outlierTest(Best.model)
```

It shows that observation 31, 16, and 20 would be influential points. However, it would be better to understand more the observations before deleting them.

Conclusions:

For the result of analysis, manual transmission performs better than automatic transmission in miles per gallon. The mean miles per gallon are significantly different between two transmissions. Also, number of cylinders, weight and horsepower are critical factors in miles per gallon. However, it would be better to include more sample size so that it would be more precise when inferring the study to bigger populations.

Appendix

Fig 1.Exploratory analysis : Relationships Between Variables.

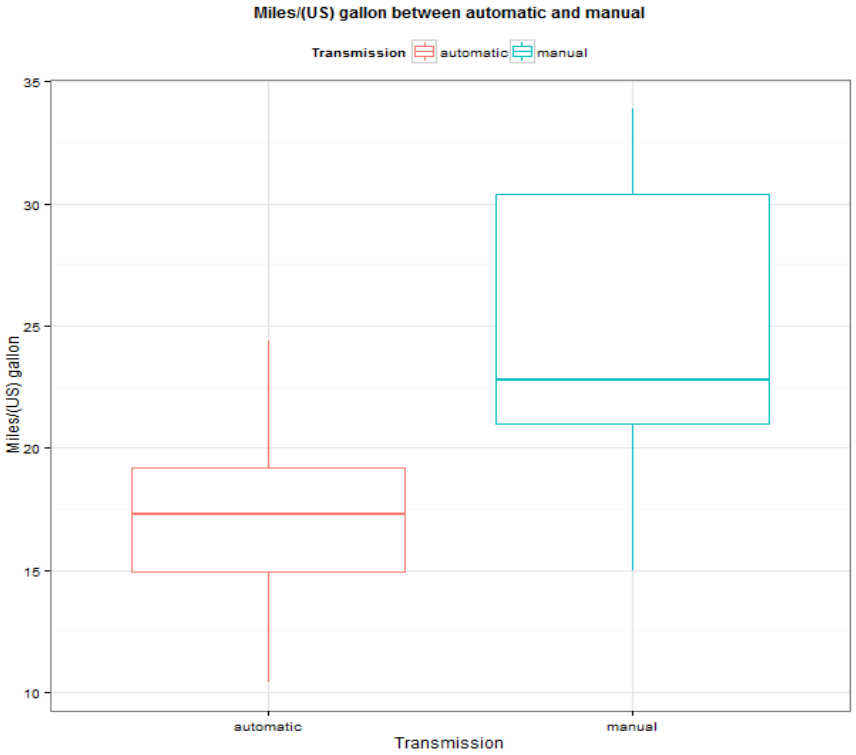

Fig 2. Miles per gallon Between Transmissions

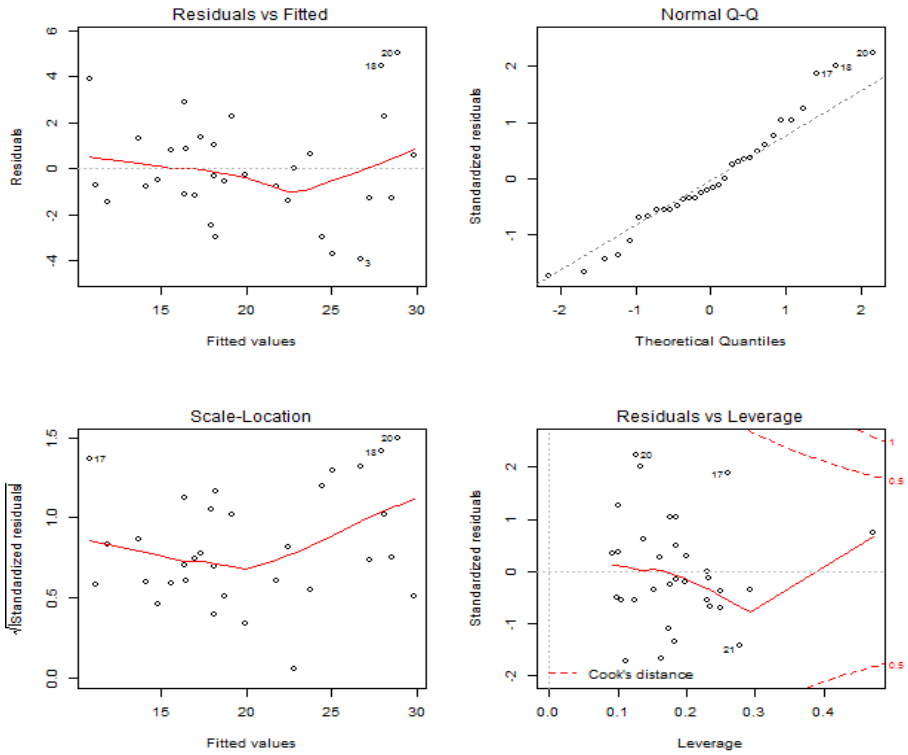

Fig 3. Regression Residual Plots of Best Model.