Regression Models Course Project

zhouyanhui

November 19, 2015

Executive Summary

This analysis is to explore the mtcars data, and try to answer the following two questions:Is an automatic or manual transmission better for MPG? Quantify the MPG difference between automatic and manual transmissions?
After linear regression analysis, there is a difference between automatic and manual transmission. Manual transmission is better than automatic for MPG. It increases 1.8 MPG swithing from automatic transmission to manual transmission with others not changed.

Read Data and Preprocess

```
library(datasets)
data("mtcars")
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am, labels = c("automatic", "manual"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Exploratory Data

Make a boxplot to show the relation between mpg and am(Transmission), appendix figure 1. According to figure 1, mpg of automatic is less than mpg of manual.

Regression Model and Model Selection

Use step function to select the best linear regression model

```
model <- lm(mpg ~ ., data = mtcars)
bestModel <- step(model, trace=FALSE)
summary(bestModel)
```

```
##
## 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
```

Compare the selected model with the base model and full Model.

```
baseModel <- lm(mpg ~ am, data = mtcars)
fullModel <- lm(mpg ~., data = mtcars)
anova(baseModel, bestModel)$P
```

```
## [1]           NA 1.688435e-08
```

```
anova(bestModel, fullModel)$P
```

```
## [1]        NA 0.9588242
```

According to the p-value, the bestModel has the most necessary variance to explain outcome. The R-square of bestModel is 0.8658799, that means it can explain 86.5879872% of the origin variance, so the bestModel is significant.

Residual Analysis

From Appendix Figure 2. The "Residuals vs Fitted" shows residuals are randomly scattered with no obvious pattern. In "Normal Q-Q", the theoretical quantiles and standardized residuals are almost in a line.

Conclusion

1. Comparing to automatic transmission, manual transmission has about 1.8 more miles per gallon.
2. MPG will decrease by about 2.5 for increasing every 1000 lb weight.
3. From 4 cylinders to 6 and 8 cylinders, MPG will decrease about 3.0 and 2.2.
4. Increasing 1 gross horsepower, MPG will only deacrease 0.032.

Appendix

Figure 1:

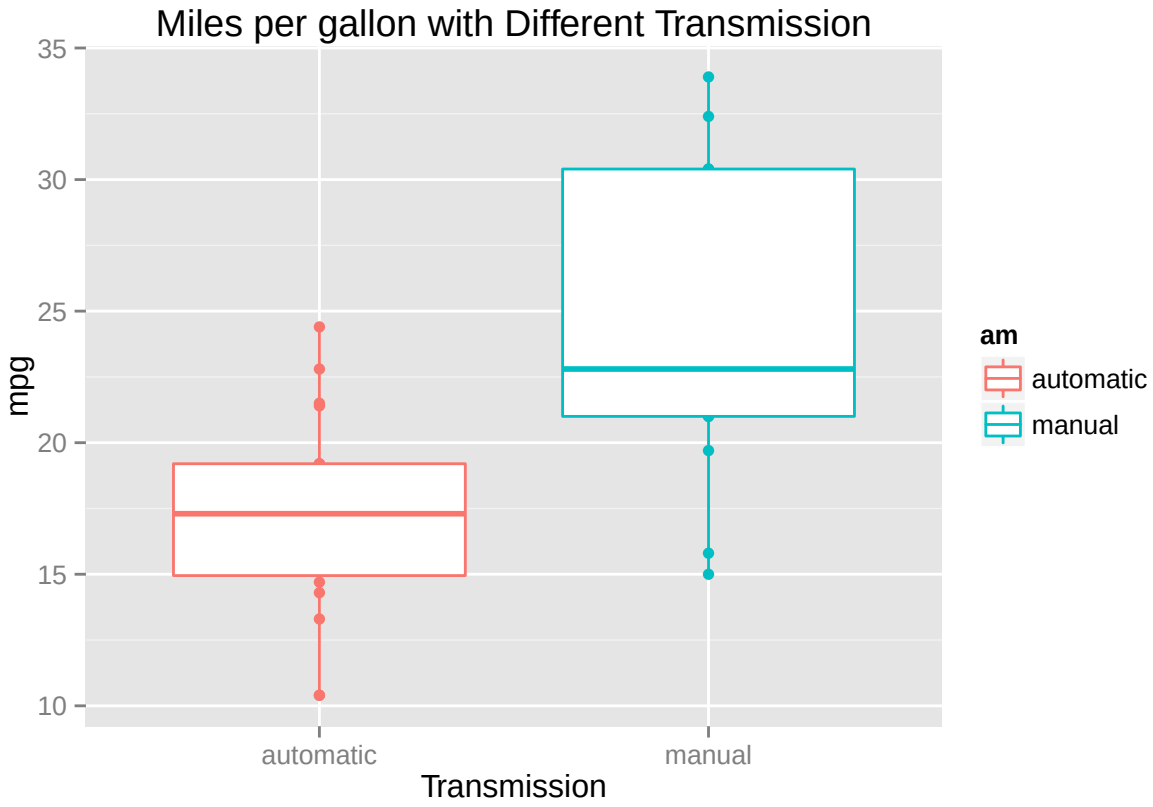
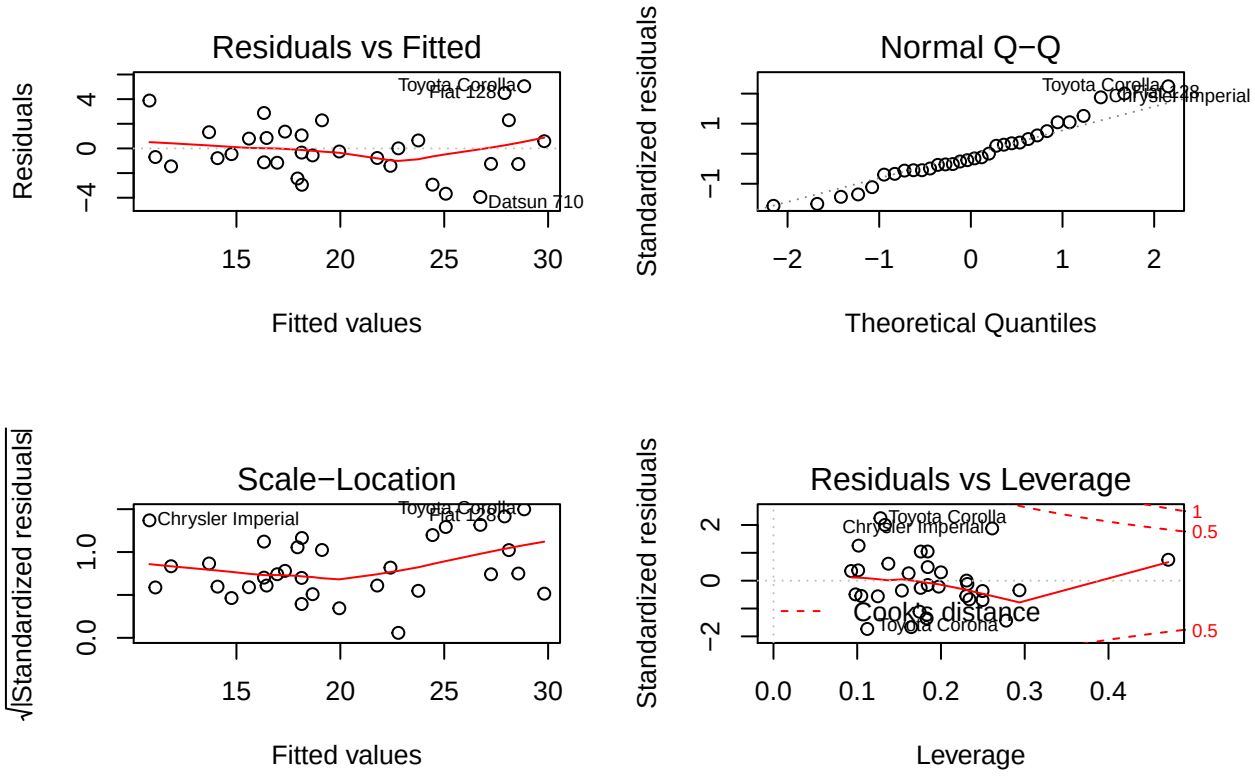


Figure 2:


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Chrysler Imperial
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Toyota Corolla
Chrysler Imperial
Toyota Corona