Regression Models Course Project

Chuwei Zhong

10/8/2019

Questions

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"

Load Data

First, we do some simple analysis for mtcars

```
data(mtcars)
```

Simple Analysis

```
cor(mtcars$mpg,mtcars[,-1])
```

```
##            cyl       disp         hp      drat         wt     qsec
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##             vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

```
#mtcars$am <- as.factor(mtcars$am)
#levels(mtcars$am) <-c("Automatic", "Manual")
```

In this analysis, cyl, disp, hp, wt and carb have negative correlaion with mpg.

automatic or manual transmission better for MPG?

The Null hypothesis: in different transmission type, the MPG has no difference.

```
t.test(mtcars$mpg~mtcars$am, conf.level=0.95)$p.value
```

```
## [1] 0.001373638
```

The p-value is 0.001373638 and very close to zero. So we reject the null hypothesis. Statistically saying, automatic transmission has a lower MPG than manual for sure.

Quantify the MPG difference between automatic and manual transmissions

To better understand the MPG difference, we deploy a multivariate linear regression.

```
linear_model <- lm(data = mtcars, mpg ~ .)
summary(linear_model)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

```
step_model = step(linear_model,trace=0,steps=10000)
summary(step_model)
```

```
##
## 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
```

Here, you can see in multivariate linear regression, 3 variables is included: wt, qsec, am and roughly explain 0.85 of the variance. Residuals(Appendix, plot 3-Residuals plot) are scattered. So, this model predict the data very well.

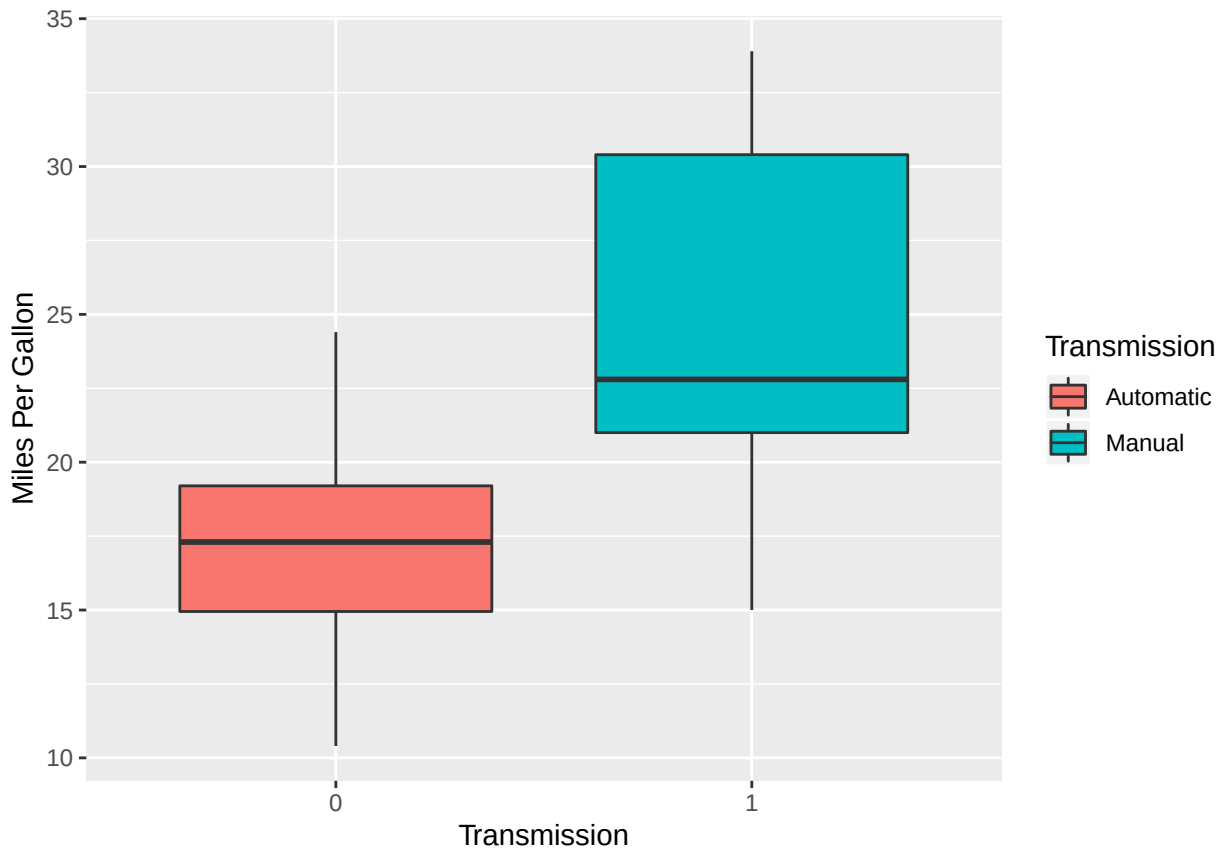
In conclusion, the influence from transmission type to MPG is not that clear. The influence is from the cooperation of transmission type, weight and acceleration speed.

Appendix

```
library(ggplot2)
```

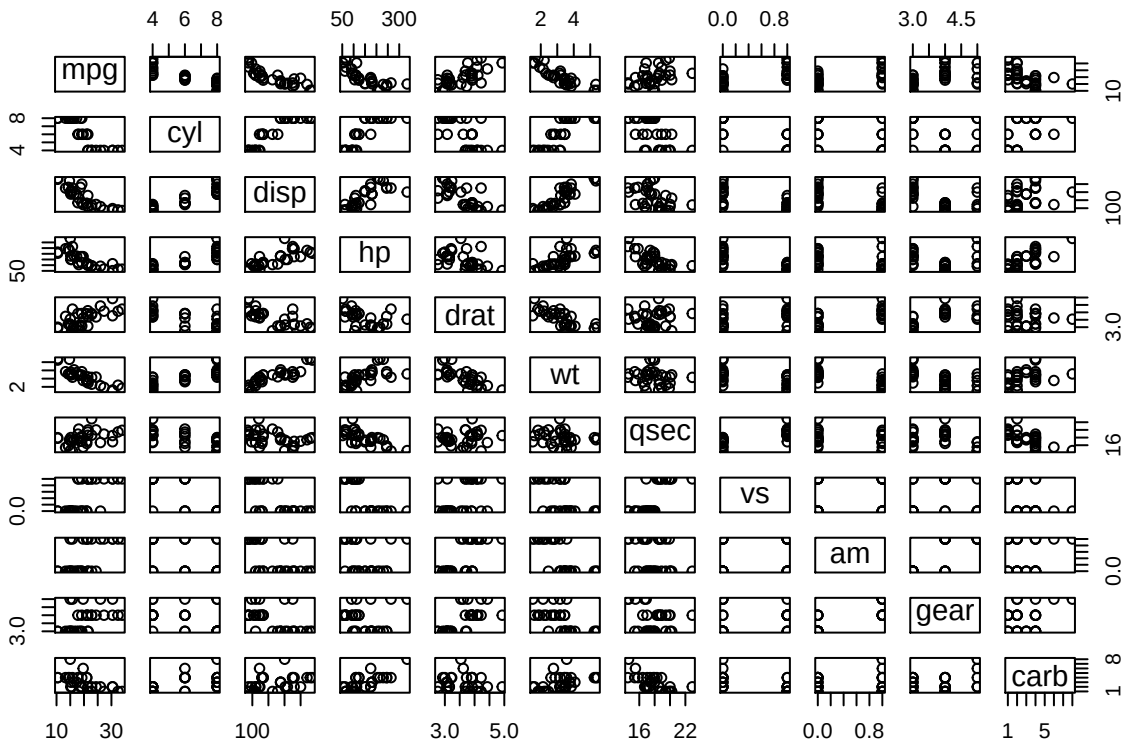
```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
ggplot(mtcars, aes(factor(am), mpg, fill = factor(am))) + geom_boxplot()+xlab("Transmission") + ylab("M
```



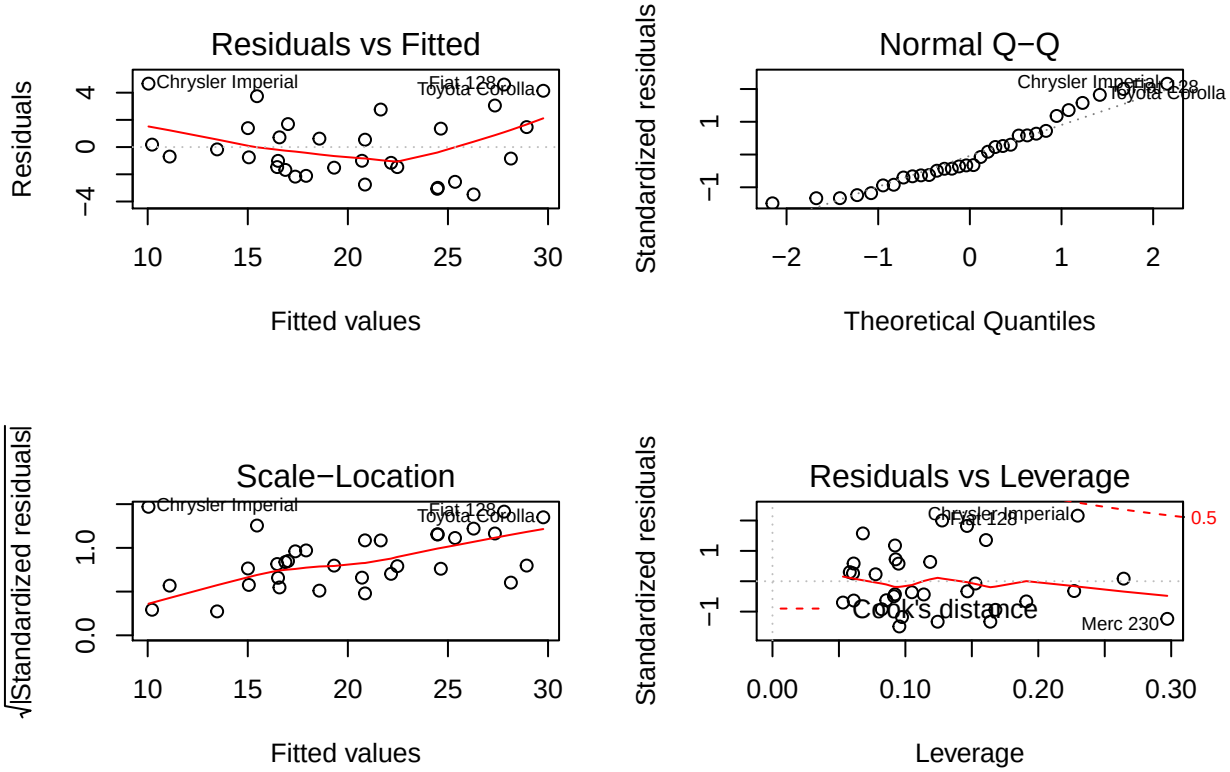
Appendix 2

```
pairs(mpg ~ ., data = mtcars)
```



Appendix 3

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


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
Cook's distance
0.5
Merc 230
Leverage