STAT0030_ICA2

Hongwei Peng

10/02/2019

R Question 1

Question 1a read the data

```
rawdata <- read.table("cars.dat", #input data
                      header=TRUE) #the first line as the names of the variables
```

Question 1b

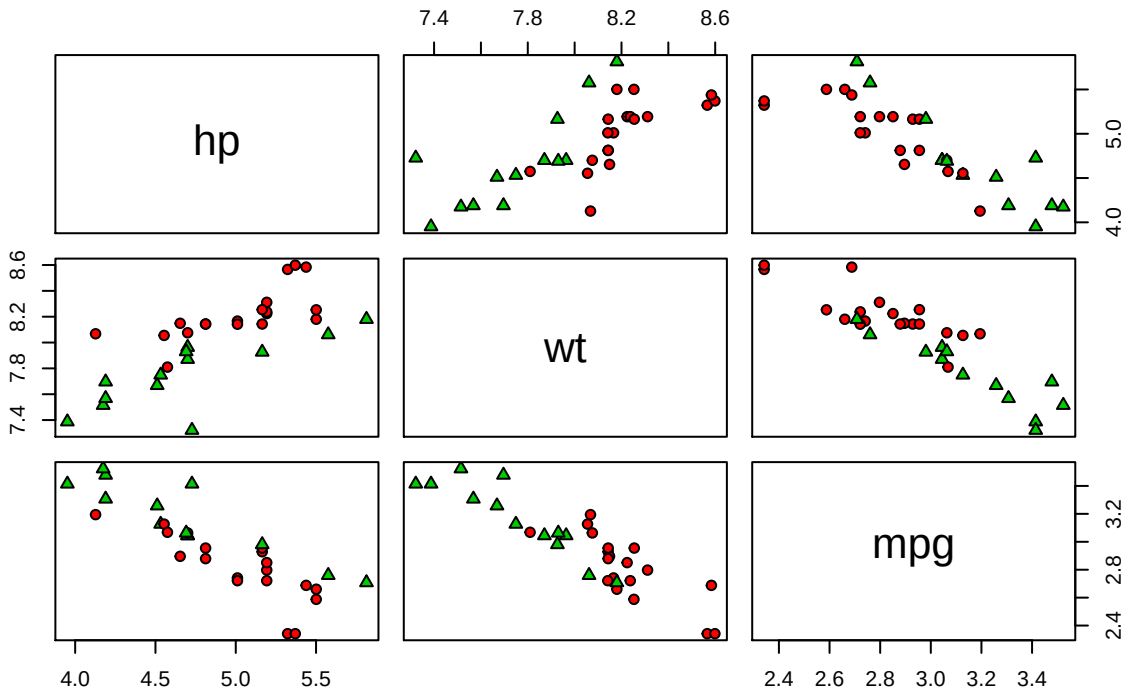
```
summary(rawdata)
```

```
##        tr               hp              wt            mpg
##  Min.   :0.0000   Min.   : 52.0   Min.   :1513   Min.   :10.40
##  1st Qu.:0.0000   1st Qu.: 96.5   1st Qu.:2581   1st Qu.:15.43
##  Median :0.0000   Median :123.0   Median :3325   Median :19.20
##  Mean   :0.4062   Mean   :146.7   Mean   :3217   Mean   :20.09
##  3rd Qu.:1.0000   3rd Qu.:180.0   3rd Qu.:3610   3rd Qu.:22.80
##  Max.   :1.0000   Max.   :335.0   Max.   :5425   Max.   :33.90
```

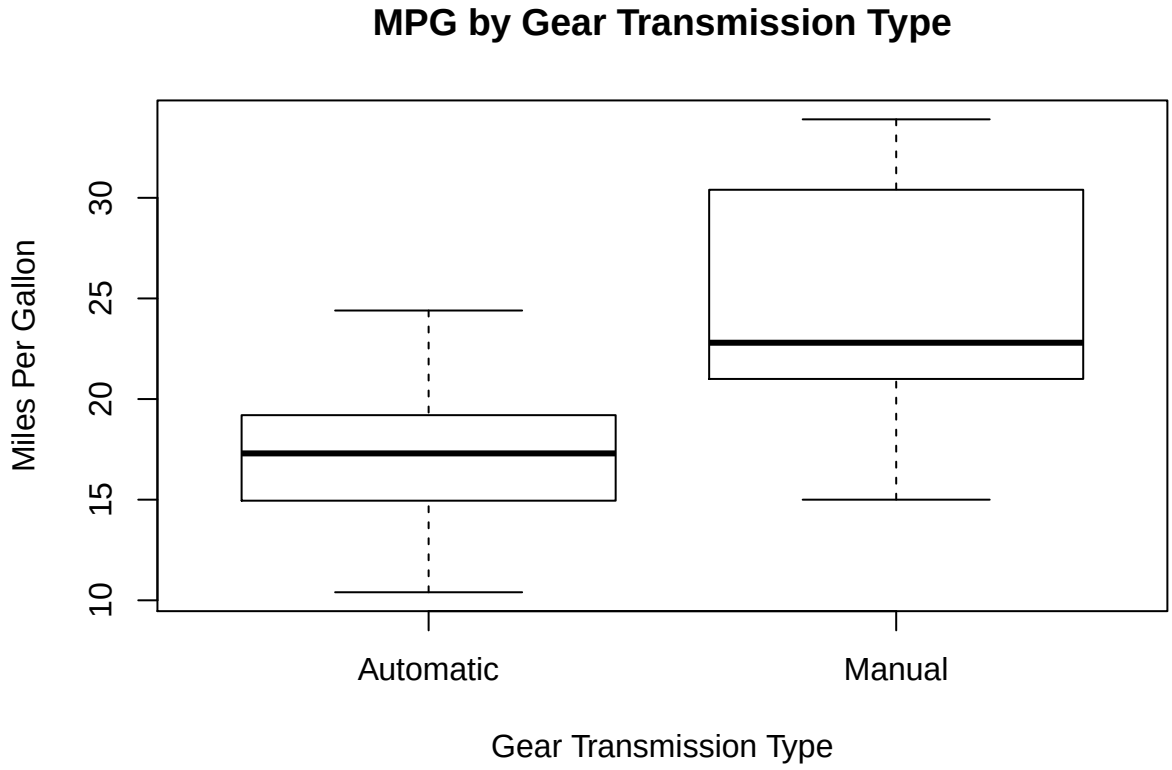
```
table(rawdata$tr)
```

```
##
##  0  1
## 19 13
```

```
logdata <- cbind(rawdata[,1],log(rawdata[,c(2,3,4)])) #log the data
names(logdata) <- c("tr","hp","wt","mpg")#rename the names of the variables
pairs(logdata[,2:4], # plot log(hp), log(wt), log(mpg)
      main = "Plot Between log Variables -- 3 species", #add the main title
      pch = c(21,24)[unclass(logdata$tr)+1], #different tr shows different shape
      bg = c("red", "green3")[unclass(logdata$tr)+1]) #different tr shows different colour
```




```
boxplot(mpg~tr, #MPG by TR
        data=rawdata, #set the dataset
        xlab="Gear Transmission Type" , #add the xlab title
        ylab="Miles Per Gallon", #add the ylab title
        main="MPG by Gear Transmission Type", #add the main title
        names=c("Automatic","Manual")) #change xlab value to character
```




```
t.test(mpg~tr, data=logdata)# the t-test above MPG is related to TR.
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by tr
## t = -3.8257, df = 23.958, p-value = 0.0008194
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5336626 -0.1596180
## sample estimates:
## mean in group 0 mean in group 1
##        2.816692        3.163332
```

question 1c

```
rawdata_model<-lm(mpg~tr+hp+wt, data=rawdata) #rawdata linear model
logdata_model<-lm(mpg~tr+hp+wt, data=logdata) #logdata linear model
summary(rawdata_model)
```

```
##
## Call:
## lm(formula = mpg ~ tr + hp + wt, data = rawdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4222 -1.7921 -0.3788  1.2250  5.5318
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.0016421  2.6423265  12.868 2.82e-13 ***
## tr           2.0841138  1.3763314   1.514 0.141169
## hp          -0.0374810  0.0096049  -3.902 0.000546 ***
## wt          -0.0028781  0.0009048  -3.181 0.003574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

```
summary(logdata_model)
```

```
##
## Call:
## lm(formula = mpg ~ tr + hp + wt, data = logdata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.204243 -0.081099 -0.003198  0.080083  0.197919
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.59990    0.86159   9.981    1e-10 ***
## tr           0.01069    0.06040   0.177 0.860813
## hp          -0.25971    0.06438  -4.034 0.000384 ***
## wt          -0.54535    0.13062  -4.175 0.000262 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1072 on 28 degrees of freedom
## Multiple R-squared:  0.883,  Adjusted R-squared:  0.8705
## F-statistic: 70.44 on 3 and 28 DF,  p-value: 3.686e-13
```

```
summary(logdata_model$residuals)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.204243 -0.081099 -0.003198  0.000000  0.080083  0.197919
```

```
best_model<-step(logdata_model, direction="both")
```

```
## Start:  AIC=-139.2
## mpg ~ tr + hp + wt
##
##        Df Sum of Sq     RSS     AIC
## - tr    1   0.00036 0.32197 -141.17
## <none>              0.32161 -139.21
## - hp    1   0.18691 0.50852 -126.54
## - wt    1   0.20023 0.52184 -125.72
##
## Step:  AIC=-141.17
## mpg ~ hp + wt
##
##        Df Sum of Sq     RSS     AIC
## <none>              0.32197 -141.17
```

```
## + tr    1   0.00036 0.32161 -139.21
## - hp    1   0.21221 0.53418 -126.97
## - wt    1   0.45943 0.78140 -114.80
```

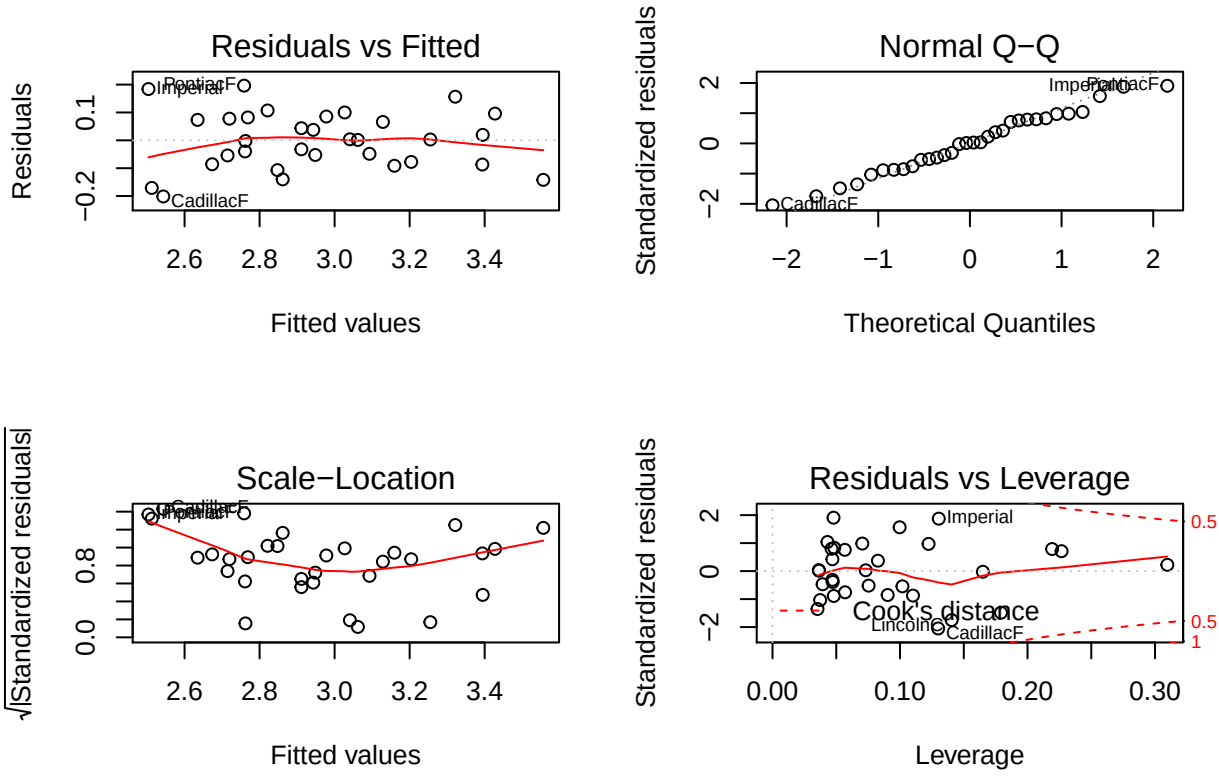
```
summary(best_model)
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = logdata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.201439 -0.079566  0.002144  0.078778  0.196144
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.71876    0.53056  16.433 3.12e-16 ***
## hp          -0.25531    0.05840  -4.372 0.000145 ***
## wt          -0.56228    0.08741  -6.433 4.89e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1054 on 29 degrees of freedom
## Multiple R-squared:  0.8829, Adjusted R-squared:  0.8748
## F-statistic: 109.3 on 2 and 29 DF,  p-value: 3.133e-14
```

```
summary(best_model$residuals)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.201439 -0.079566  0.002144  0.000000  0.078778  0.196144
```

```
par(mfrow=c(2,2)) #put 4 graphes together
plot(best_model)#plot 4 graphes as following.
```




```
#addition: if you run the last line code directly, there will be an error "figure margins too large".
#there two way to fix it: 1, click "clear all plots" and then run it.
#                         2, run >>x11() at first if you are run in Windows computer.
#BTW: there is no error if i use it in R Markdown
```