homework five

Yi (Chris) Chen

October 28, 2017

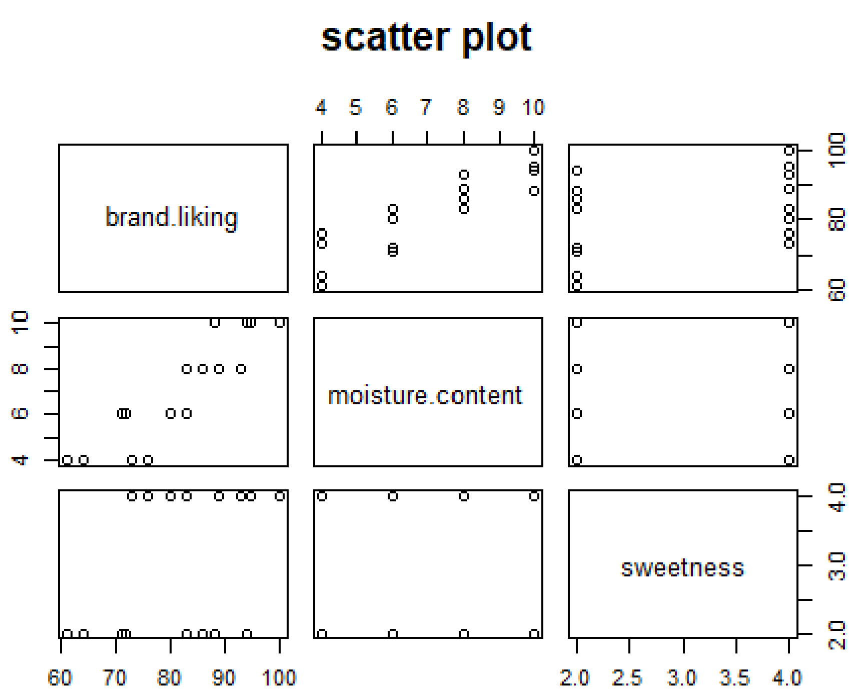
Home work five

problem one: 6.5

```
# read the data
setwd("C:/Users/cheny/Desktop/study/linear regression model/homework/homework record/homework five")
data_6.5 <- read.table('6.5.txt',header = FALSE, col.names = c('brand liking','moisture content','sweetness'))
```

(a) obtain scatter plot matrix and correlation matrix

```
# scatter plot matrix
pairs(~brand.liking+moisture.content+sweetness,data = data_6.5,main='scatter plot')
```




```
cor(data_6.5)

##                  brand.liking moisture.content sweetness
## brand.liking        1.0000000        0.8923929 0.3945807
```

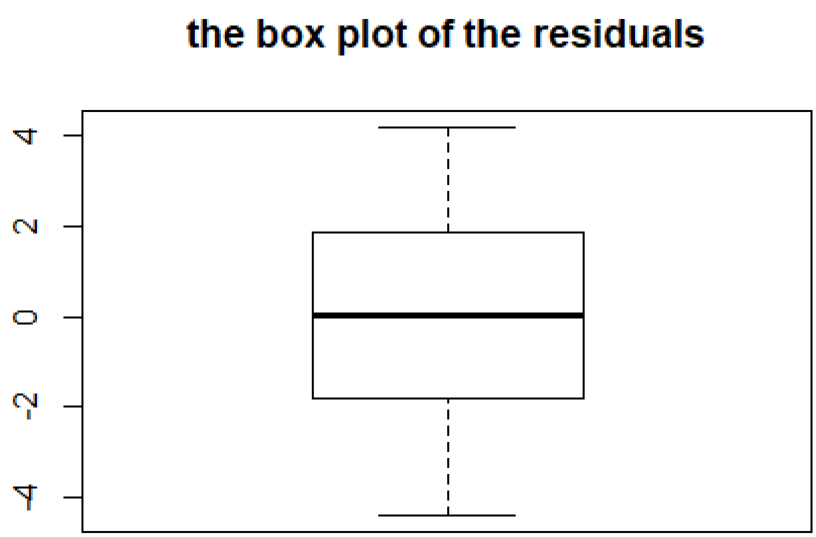
```
## moisture.content     0.8923929        1.0000000 0.0000000
## sweetness            0.3945807        0.0000000 1.0000000
```

(b) fit regression model to the data

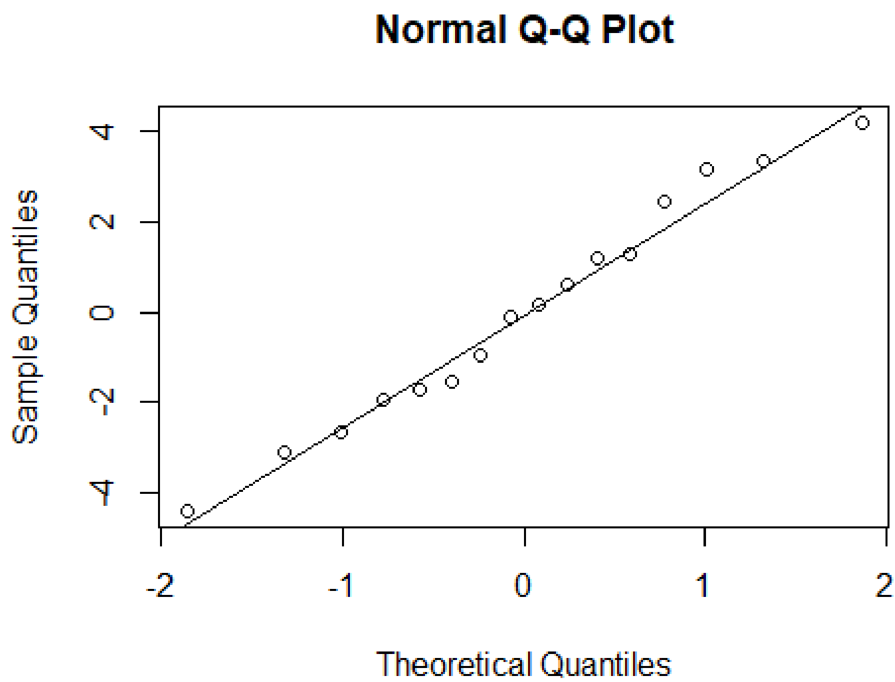
```
reg_6.5 <- lm(data = data_6.5, brand.liking ~ moisture.content + sweetness)
reg_6.5

##
## Call:
## lm(formula = brand.liking ~ moisture.content + sweetness, data = data_6.5)
##
## Coefficients:
##      (Intercept)  moisture.content         sweetness
##           37.650             4.425             4.375
```

(c) obtain the residual and draw the box plot

```
residual_6.5 <- reg_6.5$residuals
boxplot(residual_6.5,main='the box plot of the residuals')
```




```
# futher test
qqnorm(residual_6.5)
qqline(residual_6.5)
```



(f) f test for lack of fit

```
anova(lm(data = data_6.5, brand.liking ~ moisture.content + sweetness),
lm(data = data_6.5, brand.liking ~
factor(moisture.content)*factor(sweetness)))

## Analysis of Variance Table
##
## Model 1: brand.liking ~ moisture.content + sweetness
## Model 2: brand.liking ~ factor(moisture.content) * factor(sweetness)
##   Res.Df  RSS Df Sum of Sq     F Pr(>F)
## 1     13 94.3
## 2      8 57.0  5      37.3 1.047  0.453
```

problem two:6.7

(a) calculate the multipule determination R^2

```
summary(reg_6.5)

##
## Call:
## lm(formula = brand.liking ~ moisture.content + sweetness, data = data_6.5)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         37.6500     2.9961  12.566 1.20e-08 ***
## moisture.content     4.4250     0.3011  14.695 1.78e-09 ***
## sweetness            4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 on 13 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9447
## F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

(b) calculate the single determination R^2 between Yi and Yi_hat

```
Yi_hat <- reg_6.5$coefficients[1] + reg_6.5$coefficients[2] *
data_6.5$moisture.content + reg_6.5$coefficients[3] * data_6.5$sweetness
summary(lm(data_6.5$brand.liking~Yi_hat))
```

```
##
## Call:
## lm(formula = data_6.5$brand.liking ~ Yi_hat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.842e-14  4.946e+00    0.00        1
## Yi_hat       1.000e+00  5.997e-02   16.67 1.25e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.595 on 14 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9486
## F-statistic:   278 on 1 and 14 DF,  p-value: 1.246e-10
```

problem three:6.8

(a) estimate E(Yh) and iterpet interval estimation

```
X <- matrix(c(rep(1,16),data_6.5$moisture.content,data_6.5$sweetness),ncol =
3)

Y <- matrix(data_6.5$brand.liking,ncol = 1)


Xh <- matrix(c(1,5,4),ncol = 1)
beta <-
matrix(c(reg_6.5$coefficients[1],reg_6.5$coefficients[2],reg_6.5$coefficients
[3]),ncol = 1)
E_Yh <- t(Xh) %*% beta
cat('the expection of Yh is:',E_Yh)
```

```
## the expection of Yh is: 77.275

SSE <- t(Y) %*% Y - t(beta) %*% t(X) %*% Y

MSE <- SSE/(nrow(data_6.5) - 3)

S_2_b <- solve(t(X)%*%X)*as.numeric(MSE)


s_2_Yh <- t(Xh)%*%S_2_b%*%Xh
s_Yh <- sqrt(s_2_Yh)

t_value <- qt((1-0.01/2),(nrow(data_6.5)-3))

upper_value <- E_Yh + t_value * s_Yh
lower_value <- E_Yh - t_value * s_Yh

cat('the interval is:[',lower_value,',',upper_value,']')

## the interval is:[ 73.88111 , 80.66889 ]
```

(b) estimate new observation Yh(new) and predict interval

```
cat('the predict value of Yh(new) is:',E_Yh);

## the predict value of Yh(new) is: 77.275

s_2_pred <- MSE + s_2_Yh
s_pred <- sqrt(s_2_pred)

upper_value <- E_Yh + t_value * s_pred
lower_value <- E_Yh - t_value * s_pred

cat('the interval is:[',lower_value,',',upper_value,']');

## the interval is:[ 68.48077 , 86.06923 ]
```