homework five

Yi (Chris) Chen

October 28, 2017

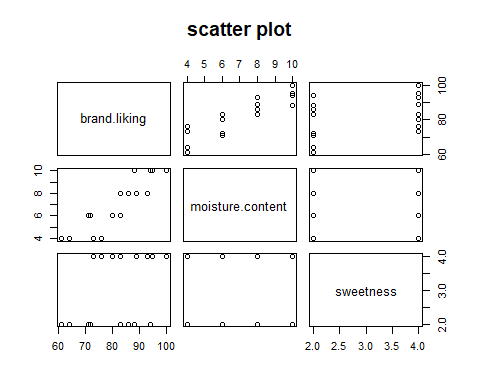
## Home work five

### 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'))

1. 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

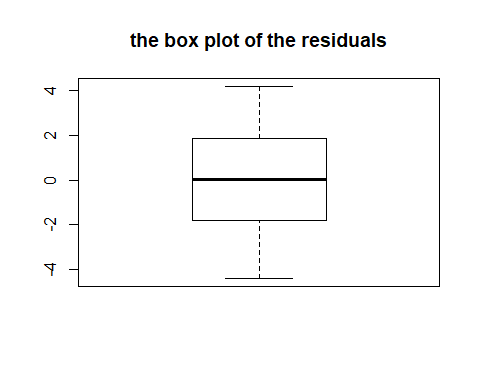
1. fit regression model to the data

reg\_6.5 <- lm(data = data\_6.5, brand.liking ~ moisture.content + sweetness)  
reg\_6.5

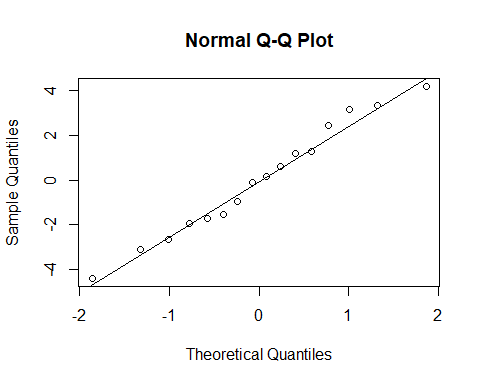
##   
## Call:  
## lm(formula = brand.liking ~ moisture.content + sweetness, data = data\_6.5)  
##   
## Coefficients:  
## (Intercept) moisture.content sweetness   
## 37.650 4.425 4.375

1. 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

library(alr3)

## Warning: package 'alr3' was built under R version 3.4.2

## Loading required package: car

## Warning: package 'car' was built under R version 3.4.2

pureErrorAnova(reg\_6.5)

## Analysis of Variance Table  
##   
## Response: brand.liking  
## Df Sum Sq Mean Sq F value Pr(>F)   
## moisture.content 1 1566.45 1566.45 219.853 4.215e-07 \*\*\*  
## sweetness 1 306.25 306.25 42.983 0.0001773 \*\*\*  
## Residuals 13 94.30 7.25   
## Lack of fit 5 37.30 7.46 1.047 0.4530065   
## Pure Error 8 57.00 7.13   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

qf(0.99,5,8)

## [1] 6.631825

### problem two:6.7

1. 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

1. 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

1. 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 ]

1. 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 ]