ST5202_Tut_2_ZHU_Xu

Name:Zhu Xu
User ID:E0337988
Student ID:A0131944H

2019/2/19

Question_1:

```
data<-read.table(
  '/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/copier_maintenance.txt')
X<-data[,2]
Y<-data[,1]
n=length(Y)
reg<-lm(Y~X)
newobs=data.frame(Y=c(NA),X=c(6))
CI=predict(reg, newdata = newobs, interval = 'confidence', level=0.90)
PL=predict(reg, newdata = newobs, interval = 'predict', level=0.90)
```

a)

```
summary(CI)
```

```
##       fit             lwr             upr
##  Min.   :89.63   Min.   :87.28   Min.   :91.98
##  1st Qu.:89.63   1st Qu.:87.28   1st Qu.:91.98
##  Median :89.63   Median :87.28   Median :91.98
##  Mean   :89.63   Mean   :87.28   Mean   :91.98
##  3rd Qu.:89.63   3rd Qu.:87.28   3rd Qu.:91.98
##  Max.   :89.63   Max.   :87.28   Max.   :91.98
```

b)

```
summary(PL)
```

```
##       fit             lwr             upr
##  Min.   :89.63   Min.   :74.46   Min.   :104.8
##  1st Qu.:89.63   1st Qu.:74.46   1st Qu.:104.8
##  Median :89.63   Median :74.46   Median :104.8
##  Mean   :89.63   Mean   :74.46   Mean   :104.8
##  3rd Qu.:89.63   3rd Qu.:74.46   3rd Qu.:104.8
##  Max.   :89.63   Max.   :74.46   Max.   :104.8
```

Question_2:

a)

```
summary(reg)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## X            15.0352     0.4831  31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

```
t_value = 31.123
s_b1 = 0.4831
t_value < qt((1-0.01/2),(n-2))
```

```
## [1] FALSE
```

```
#reject H_0
```

b)

```
t_value < qt((1-0.01),(n-2))
```

```
## [1] FALSE
```

```
#reject H_0
```

c)

```
U = reg$coefficients[2]+qt((1-0.01/2),(n-2))*s_b1
L = reg$coefficients[2]-qt((1-0.01/2),(n-2))*s_b1
c(L,U)
```

```
##        X        X
## 13.73324 16.33725
```

Question_3:

a)

$$\text{Full Model: } Y = \beta_0 + \beta_1 X + \epsilon$$

$$\text{Reduced Model: } Y = \beta_0 + \epsilon$$

b)

```
anova(reg)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X          1  76960   76960  968.66 < 2.2e-16 ***
## Residuals 43   3416      79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#SSE(F)=3416
#SSE(R)=76960
#dfF=43
#dfR=1
F=968.66
F < qf(0.99,1,(n-2))
```

```
## [1] FALSE
```

```
p_value = 1-pf(F,1,(n-2))
p_value
```

```
## [1] 0
```

c)

$$qT(1 - \tfrac{0.01}{2})^2 = qF(0.99, 1, (n-2))$$

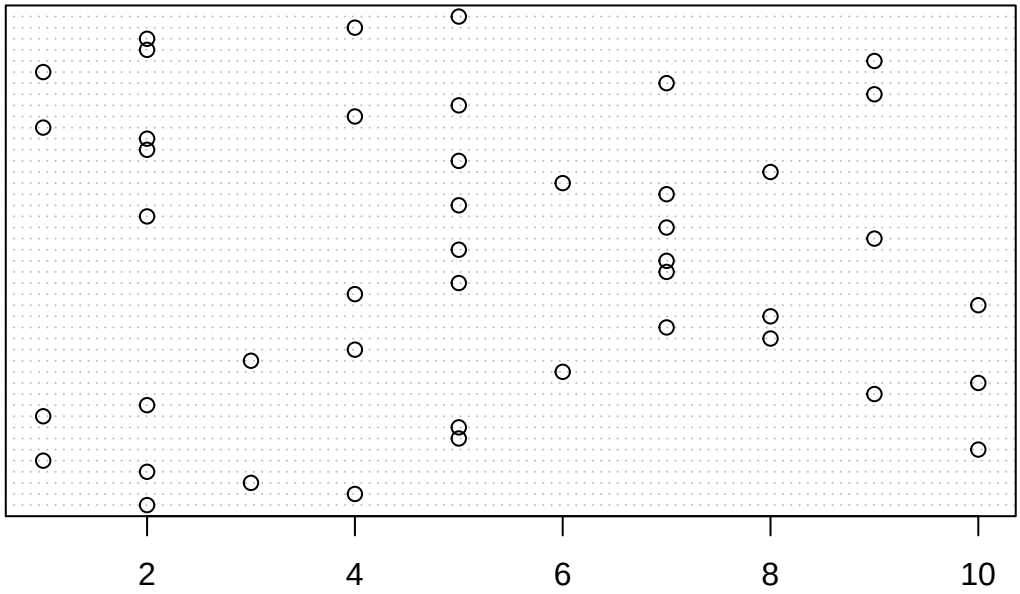
$$(\tfrac{b_1}{s\{b_1\}})^2 = \tfrac{SSR}{MSE}$$

$$\Rightarrow \quad equivalent$$

Question_4:

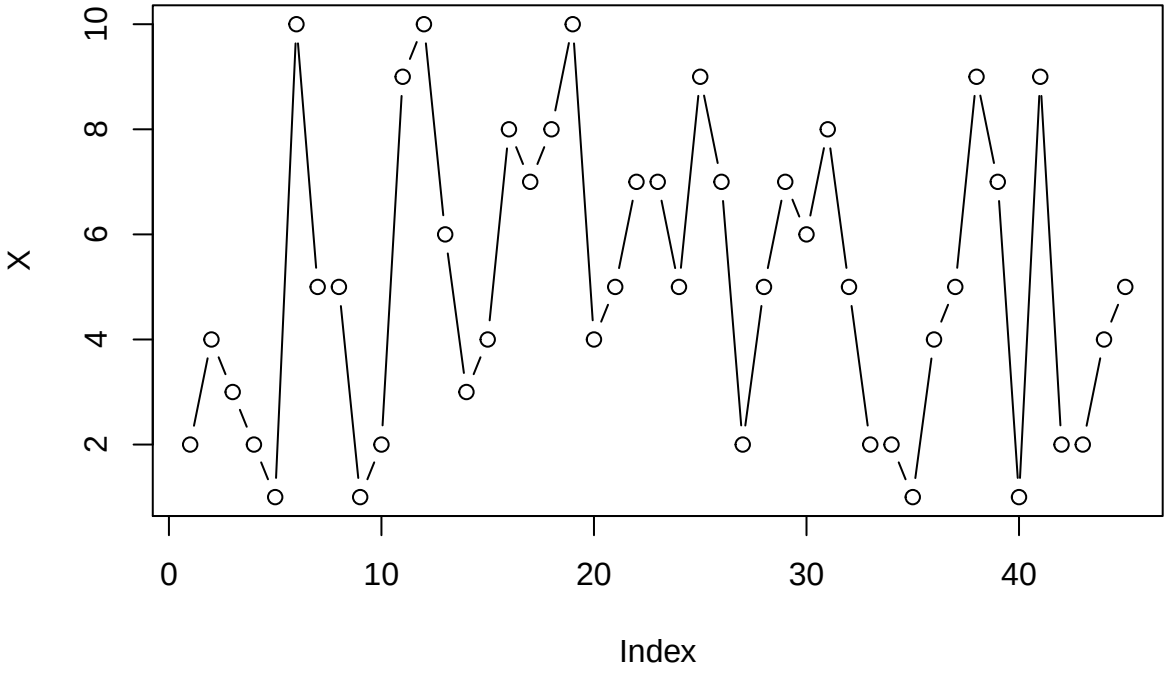
a)

```
dotchart(X)
```



b)

```
plot(X,type = 'b')
```



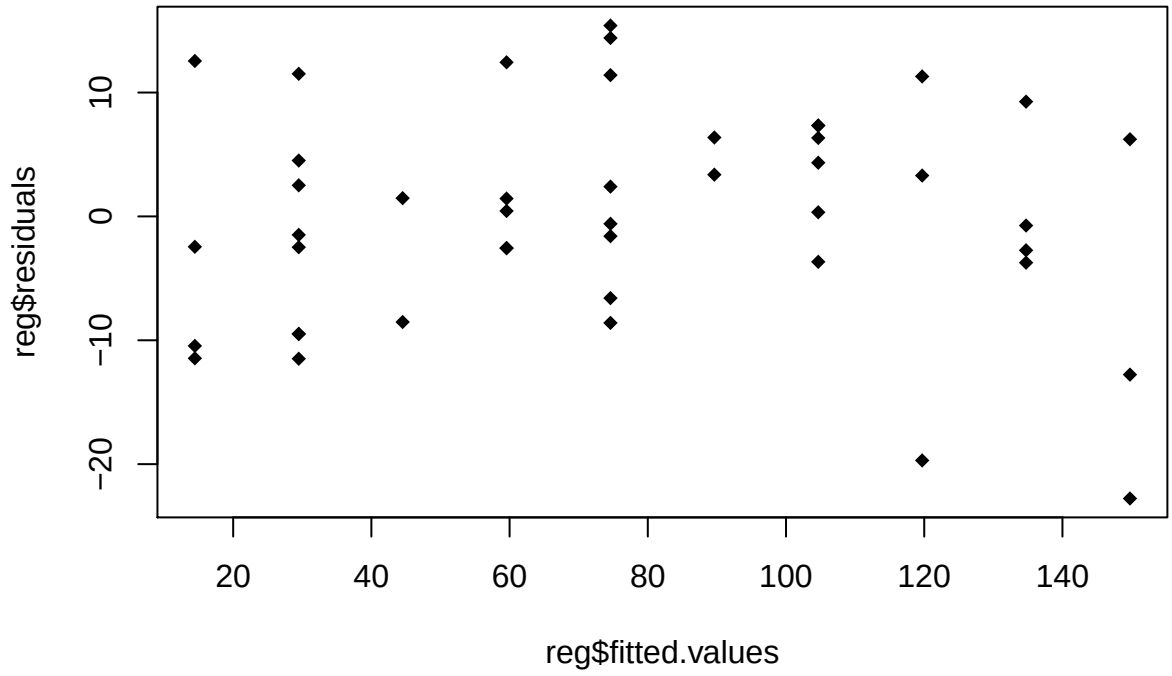
c)

```
stem(reg$residuals)
```

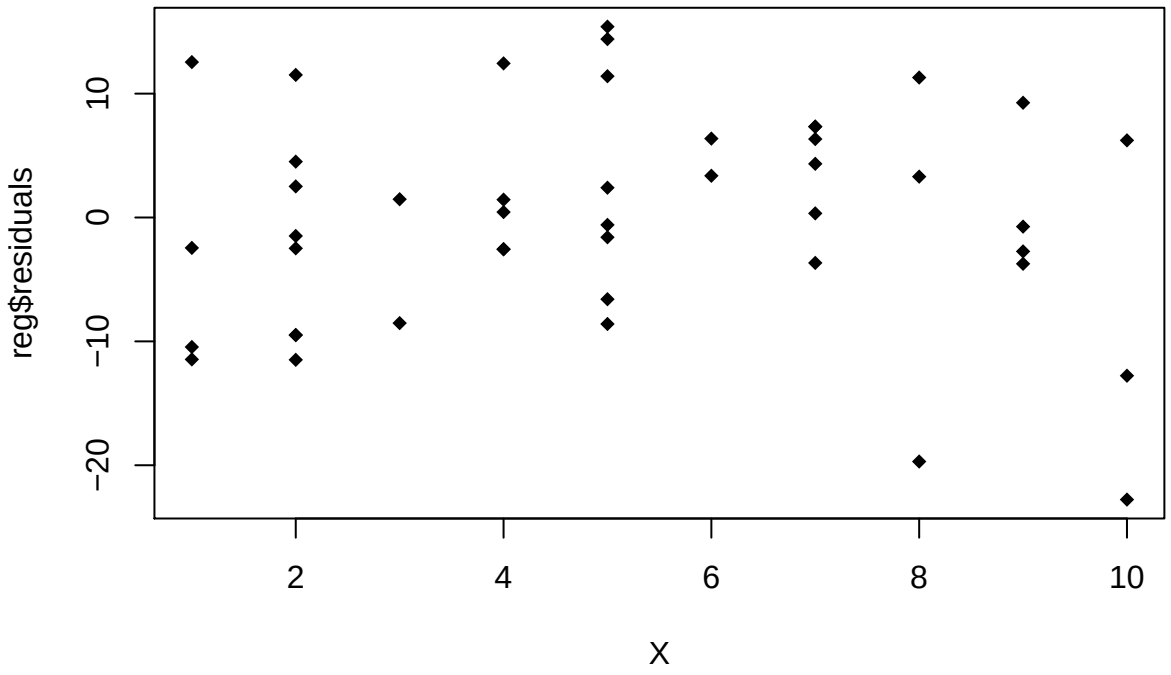
```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   -2 | 30
##   -1 |
##   -1 | 3110
##   -0 | 99997
##   -0 | 44333222111
##    0 | 001123334
##    0 | 5666779
##    1 | 112234
##    1 | 5
```

d)

```
plot(reg$fitted.values,reg$residuals,pch=18)
```

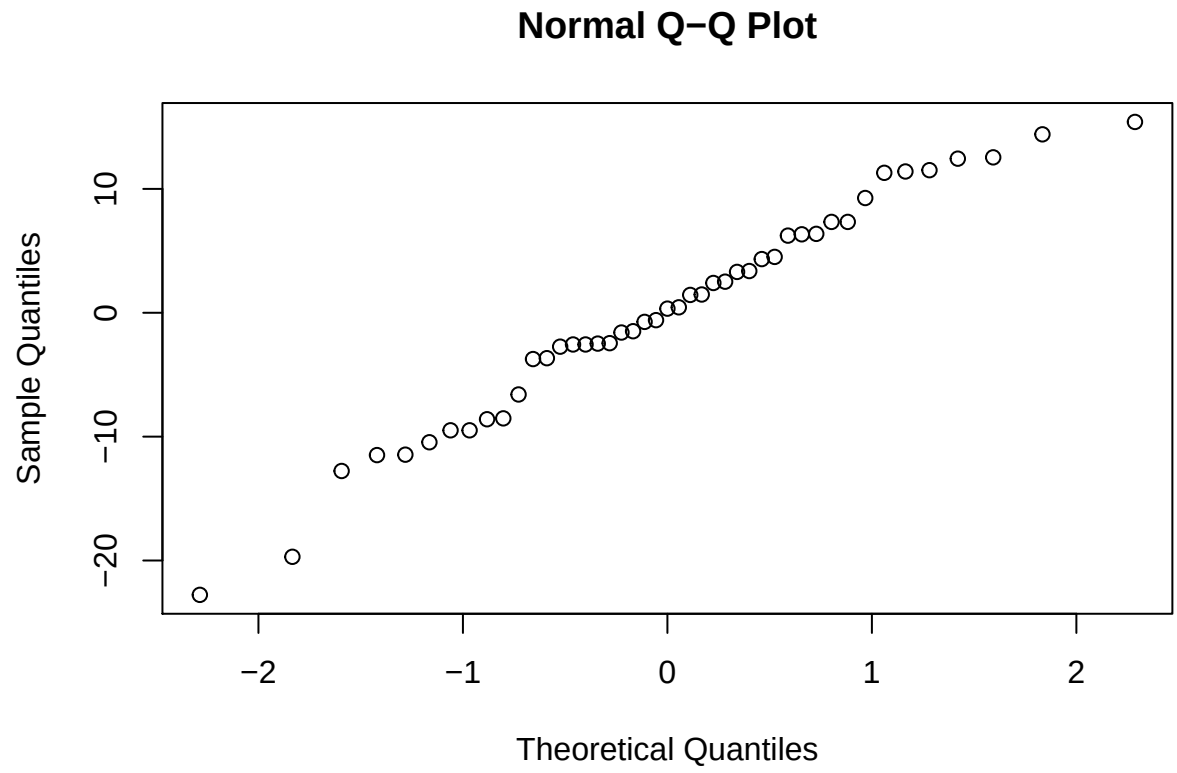



```
plot(X,reg$residuals,pch=18)
```



e)

```
qqnorm(reg$residuals)
cc=cor(qqnorm(reg$residuals)$x,qqnorm(reg$residuals)$y)
```

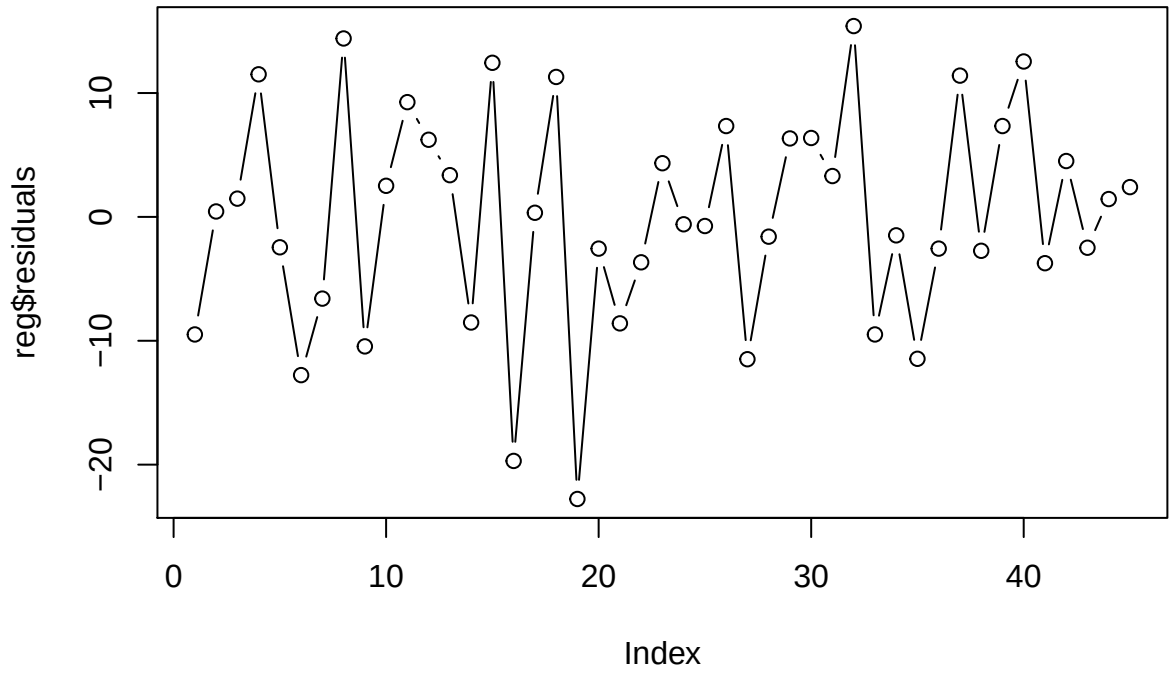



```
cc #cc>critical value
```

```
## [1] 0.9889098
```

f)

```
plot(reg$residuals,type = 'b')
```



g)

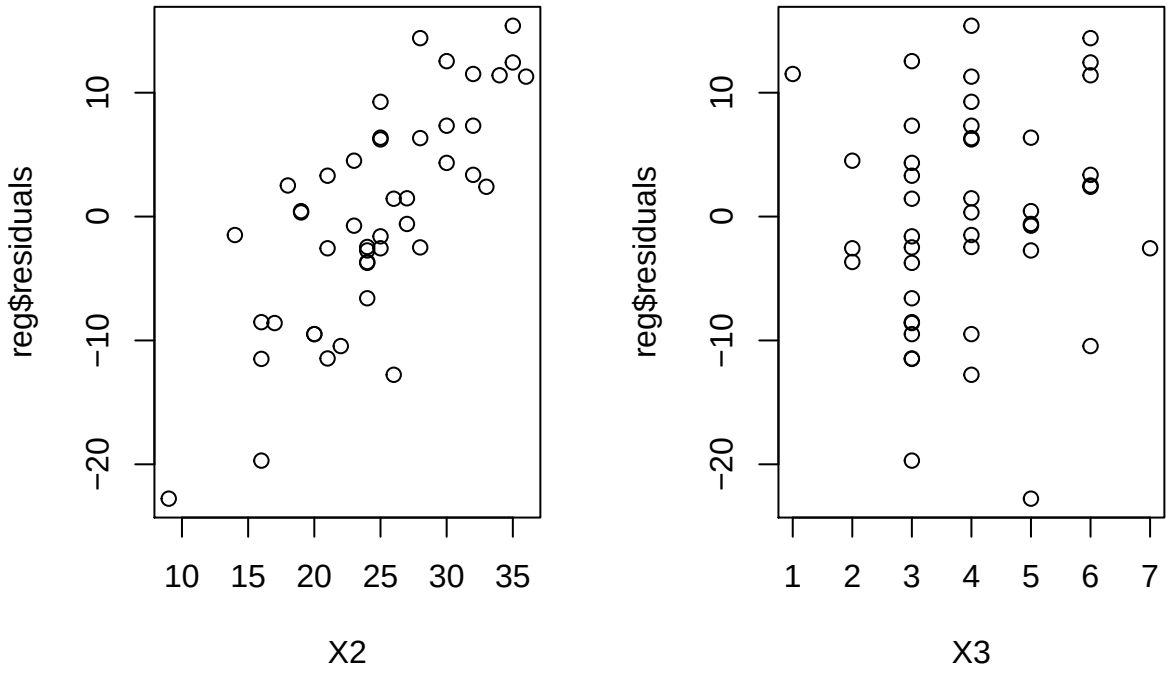
```
#BFtest
index=which(X<=median(X))
e1=reg$residuals[index]
e2=reg$residuals[-index]
d1=e1-median(e1)
d2=e2-median(e2)
n1=length(d1)
n2=length(d2)
temp1=sum((d1-mean(d1))^2)
temp2=sum((d2-mean(d2))^2)
s=sqrt((temp1+temp2)/(n-2))
tBF=(mean(d1)-mean(d2))/sqrt(1/n1+1/n2)/s
tBF<qt((1-0.05/2),(n-2))
```

```
## [1] TRUE
```

```
#CONSTANT
```

h)

```
newdata<-read.table(
  '/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/q3_4.txt')
X2=newdata[,3]
X3=newdata[,4]
par(mfrow=c(1,2))
plot(X2,reg$residuals)
plot(X3,reg$residuals)
```



Quetion_5:

a+b)

```
SC<-read.table(
  '/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/solution_concentration.txt')
y = SC[,1]
x = SC[,2]
reg1 = lm(y~x)
colnames(SC) = c('y','x')
full.model = lm(y~factor(x),data = SC)
reduced.model = reg1
anova(reduced.model,full.model)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ factor(x)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     13 2.9247
## 2     10 0.1574  3    2.7673 58.603 1.194e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
F_statistics = 58.603
F_statistics > qf((1-0.025),3,length(x)-5)
```

```
## [1] TRUE
```

```
#reject H_0
```

c)

```
#SSE(F)=0.1574
#SSE(R)=2.9247
#dfF=10
#dfR=13
F_statistics=58.603
p_value1 = 1-pf(F_statistics,3,length(x)-5)
p_value1
```

```
## [1] 1.194517e-06
```

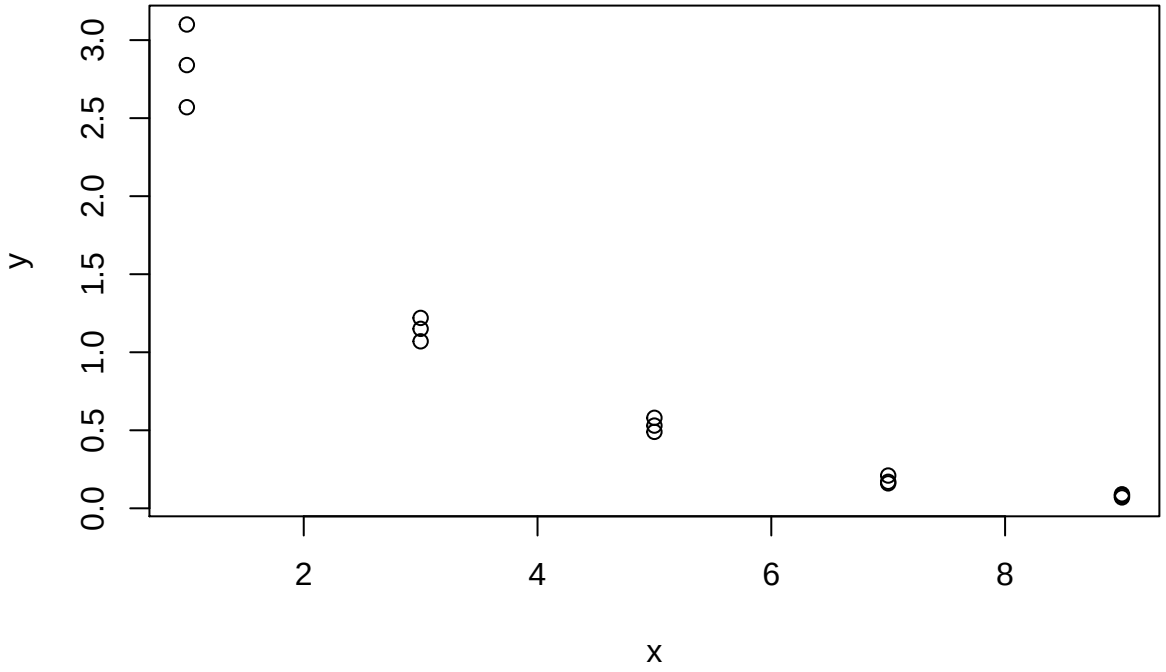
d+e)

The test in b is not enough to judge, we also need further study of residuals.

Question_6:

a)

```
par(mfrow=c(1,1))
plot(x,y)
```




```
#Y'=log10Y is approriate.
```

c)

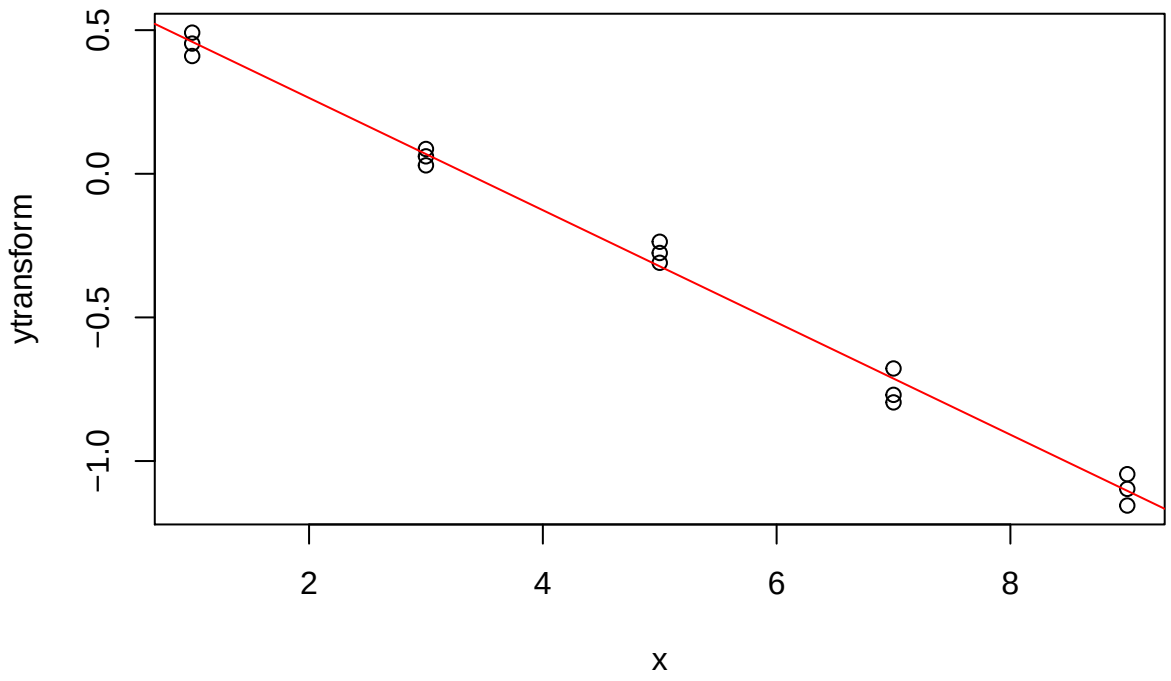
```
ytransform=log10(y)
reg2=lm(ytransform~x)
summary(reg2)
```

```
##
## Call:
## lm(formula = ytransform ~ x)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.082958 -0.044421  0.006813  0.033512  0.085550
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.654880   0.026181   25.01 2.22e-12 ***
## x           -0.195400   0.004557  -42.88 2.19e-15 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04992 on 13 degrees of freedom
## Multiple R-squared:  0.993,  Adjusted R-squared:  0.9924
## F-statistic:  1838 on 1 and 13 DF,  p-value: 2.188e-15
```

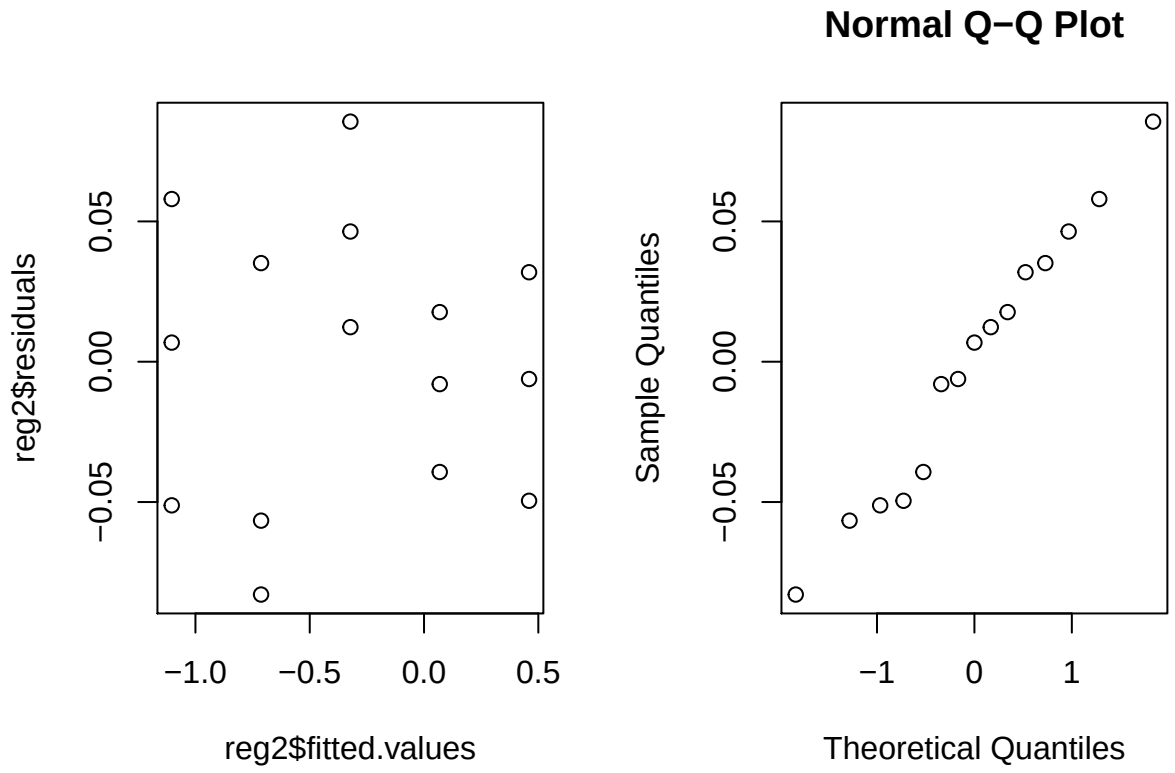
d)

```
b0<-reg2$coefficients[1]
b1<-reg2$coefficients[2]
plot(x,ytransform)
abline(b0,b1,col='red')
```



e)

```
par(mfrow=c(1,2))
plot(reg2$fitted.values,reg2$residuals)
qqnorm(reg2$residuals)
```



f)

```
b1
```

```
##          x
## -0.1954003
```

```
b0
```

```
## (Intercept)
##   0.6548798
```

$$log(y) = 0.6548798 - 0.1954003x$$

$$y = exp\{0.6548798 - 0.1954003x\}$$

$$y = 4.517309exp\{-0.1954003x\}$$