In the name of God

Producer:
Mehrab Atighi

Subject:

Exercise chapter thiry

Date:
4/4/2021

Supervisor:
Dr.Seyed Nourollah Mousavi

solve 5 bottom questions:

1)

a) According to the Bottem data make main model woth 18 points and get summary function?

```
#Exercise one:
rm(list=ls())
#intruducing Data:
x<-c(1.5,1.7,2,2.2,2.5,2.5,2.7,2.9,3,3.5,3.8,4.2,4.3,4.6,4,5.1,5.2,5.5)
y<-c(1,2.5,3.5,3,3.1,3.6,2.2,3.9,4,4,4.2,4.1,4.8,1.2,5.1,5.1,4.8,5.3)

A=3.4;B=9.5;C=9.5
a=8;b=8;c=2.5
#Make Models and get needed information:
#Make First Regression Model(with main Data)
#a)
fit1<-lm(y~x)
#Get needed information from this Model
summary(fit1)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1998 -0.0290  0.3070  0.5749  1.0834
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.4616     0.7204   2.029  0.05945 .
## x             0.6387     0.1996   3.200  0.00558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.025 on 16 degrees of freedom
## Multiple R-squared:  0.3902, Adjusted R-squared:  0.3521
## F-statistic: 10.24 on 1 and 16 DF,  p-value: 0.005578
```

now we can see that or p_value for x is lower than 0.05 and the betha0 = 1.4616,

betha1 = 0.6387, the Residual standard error = 1.025,

R-squared = 0.3902,t(betha1) = 3.200

b) According to the Bottem data make main model woth 18 points+A&a and get summary function?

```
#b)
#make Second regssion Model(With A point & main Data)
x[19]= A ; y[19]= a
fit2<-lm(y~x)
#Get needed information from this Model
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4296 -0.2435  0.0813  0.3591  4.1368
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.6914     1.0036   1.685   0.1102
## x             0.6387     0.2789   2.290   0.0351 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.432 on 17 degrees of freedom
## Multiple R-squared:  0.2358, Adjusted R-squared:  0.1908
## F-statistic: 5.245 on 1 and 17 DF,  p-value: 0.03506
```

now we can see that or p_value for x is lower than 0.05 and the betha0 = 1.6914,

betha1 = 0.6387, the Residual standard error = 1.432,

R-squared = 0.2358,t(betha1) = 2.290.

c) According to the Bottem data make main model woth 18 points+B&b and get summary function?

```
#c)
#make third regssion Model(With B point & main Data)
```

```
x[19]= B ; y[19]= b
fit3<-lm(y~x)
#Get needed information from this Model
summary(fit3)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2633 -0.0281  0.2221  0.5561  1.0464
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.3223     0.5251   2.518   0.0221 *
## x             0.6828     0.1270   5.375 5.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9973 on 17 degrees of freedom
## Multiple R-squared:  0.6296, Adjusted R-squared:  0.6078
## F-statistic:  28.9 on 1 and 17 DF,  p-value: 5.034e-05
```

now we can see that or p_value for x is lower than 0.05 and the betha0 = 1.3223,

betha1 = 0.6828, the Residual standard error = 0.9973,

R-squared = 0.6296,t(betha1)= 5.375.

d) According to the Bottem data make main model woth 18 points+C&c and get summary function?

```
#d)
#make forth regssion Model(With C point & main Data)
x[19]= C ; y[19]= c
fit4<-lm(y~x)
#Get needed information from this Model
summary(fit4)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5206 -0.5277  0.4463  0.7961  1.4797
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.9518     0.6646   4.442 0.000358 ***
## x             0.1671     0.1608   1.040 0.313055
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.262 on 17 degrees of freedom
## Multiple R-squared:  0.05978,    Adjusted R-squared:  0.004475
## F-statistic: 1.081 on 1 and 17 DF,  p-value: 0.3131
```

now we can see that or p_value for x is not lower than 0.05 and the betha0 = 2.9518,

betha1 = 0.1671, the Residual standard error = 1.262,

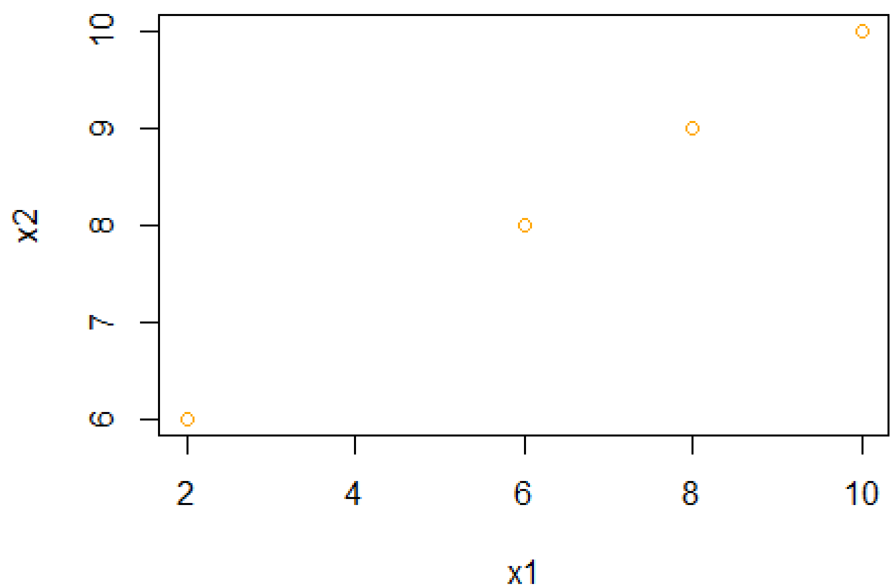
R-squared = 0.05978,t(betha1)=1.040

End.

---

2) show that the bottem datas variables (x1,x2) have linear relation.

```r
#Exercise Two:
x1<-c(2,8,6,10)
x2<-c(6,9,8,10)
#produce variables plot:
par(mfrow=c(1,1))
plot(x1,x2,col="orange")
```



We can see a possetive relation between X1 & X2, so we can say when they have a linear relation. it will be possible to make a mistake make a regression model. End.

---

3) This question involves the use of simple linear regression on the Auto data set.

a) Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Use the summary() function to print the results. Comment on the output. For example:

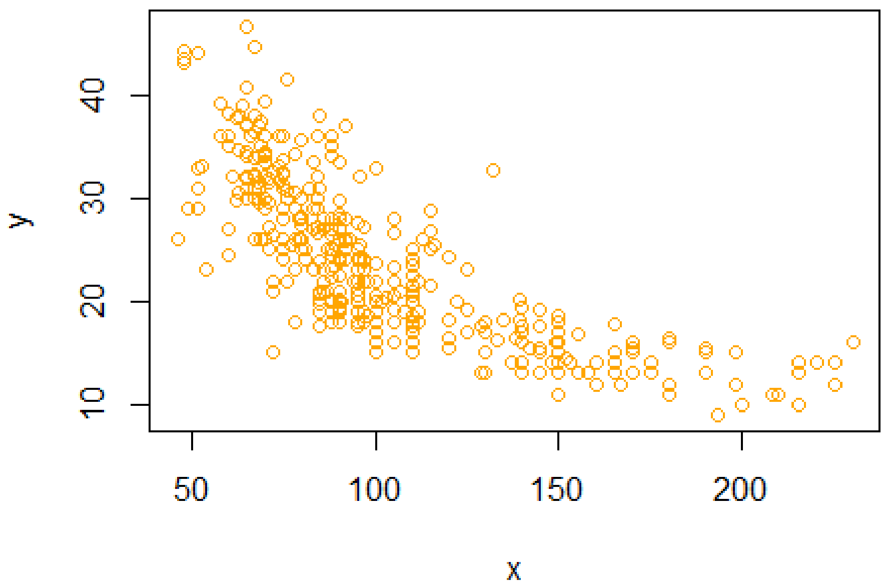
```
#Exercise Third(eigth of Book):
#library Data:
library("ISLR")

## Warning: package 'ISLR' was built under R version 4.0.3

y<-Auto$mpg
x<-Auto$horsepower
#a)
```

```r
fit1<-lm(y~x)
summary(fit1)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5710  -3.2592  -0.3435   2.7630  16.9240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.935861   0.717499   55.66   <2e-16 ***
## x           -0.157845   0.006446  -24.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.906 on 390 degrees of freedom
## Multiple R-squared:  0.6059, Adjusted R-squared:  0.6049
## F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

i. Is there a relationship between the predictor and the response?

```r
#i)
plot(x, y, col = "orange", type = "p")
```



We can see a nagative relation between X & Y its fixed for x > 150 .

ii. How strong is the relationship between the predictor and the response?

```
#ii
summary(fit1)[8]

## $r.squared
## [1] 0.6059483

anova(fit1)[5]

##              Pr(>F)
## x         < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

the r.squared is equal to 0.6059483.

the p value of the Anova of our model is very lower than 0.05.

So we can say that this linear regression model is significant and the r. squared show that its not very strong relation!.

iii. Is the relationship between the predictor and the response positive or negative?

```
#iii)
cor(x,y)
```

```
## [1] -0.7784268
```

the correlation value is equal to -0.7784268 , so we have a nagative relation between X & Y

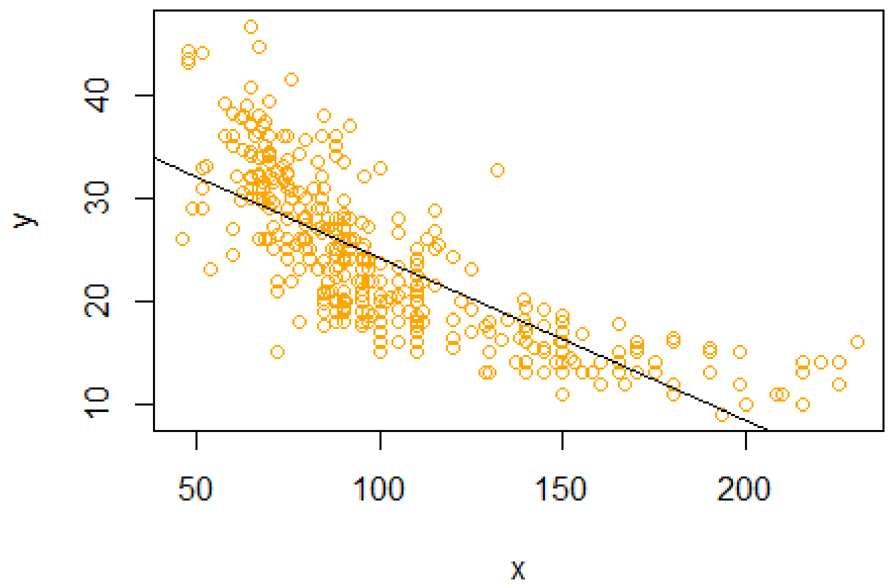
iv. What is the predicted mpg associated with a horsepower of 98? What are the associated 95 % confidence and prediction intervals?

```
#iv)
#make a f(x) function and calculate f(98):
f<-function(X){
  f=as.numeric(fit1$coefficients[1])+ as.numeric((X*fit1$coefficients[2]))
  return(f)
}

f(98)

## [1] 24.46708

#predictioin confidence interval 95% for f(98) predict:
predict(fit1,newdata = as.data.frame(x<-c(98)),interval = "confidence")

##        fit      lwr      upr
## 1 24.46708 23.97308 24.96108
```

clearly we can see the lower and upper limit of our confidence interval and the prediction value.

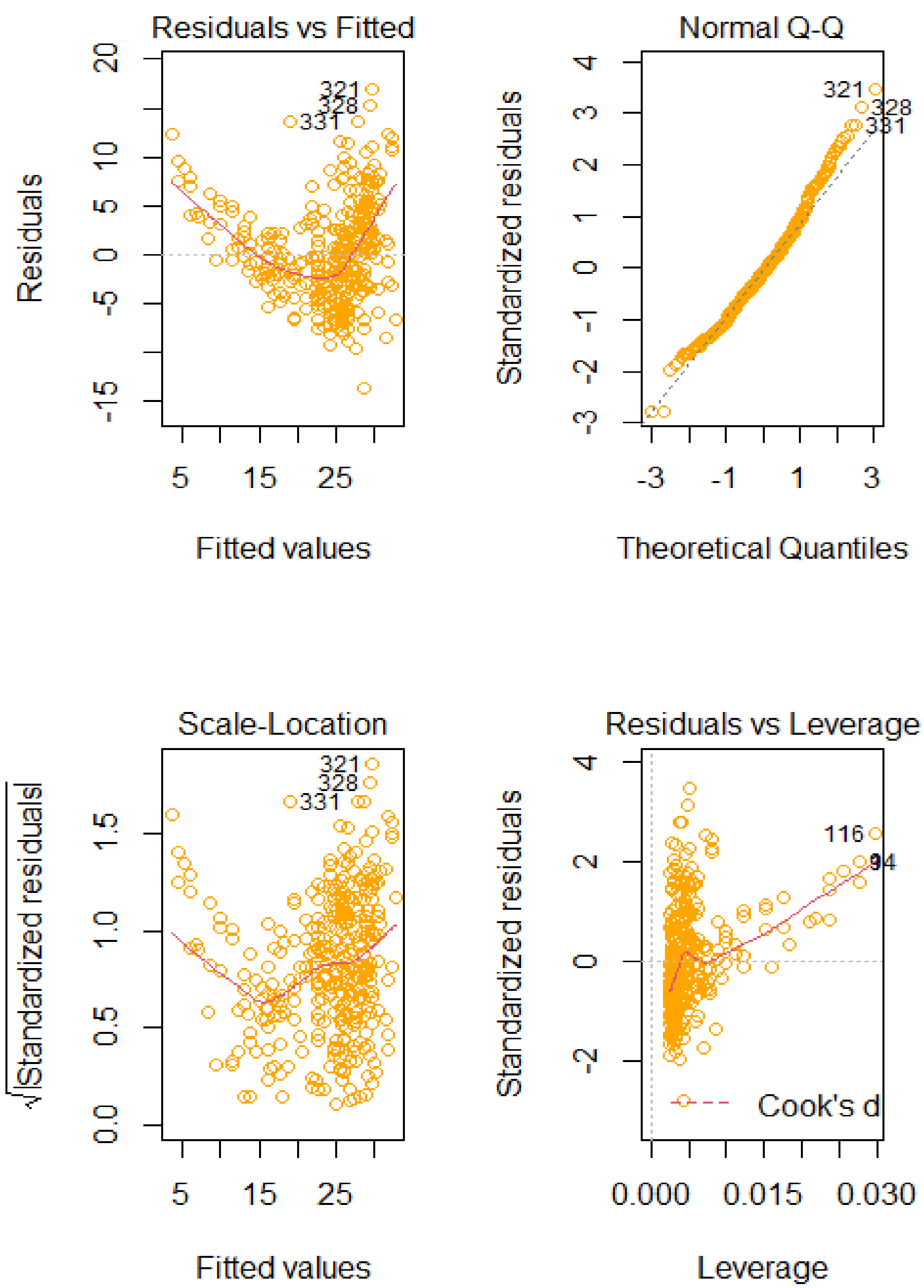
b) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

```
#b)
x<-Auto$horsepower
y<-Auto$mpg
plot(x, y, col = "orange", type = "p")
abline(fit1,col="black")
```



c) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.

```
#c)
par(mfrow= c(1,2))
plot(fit1,col="orange")
```



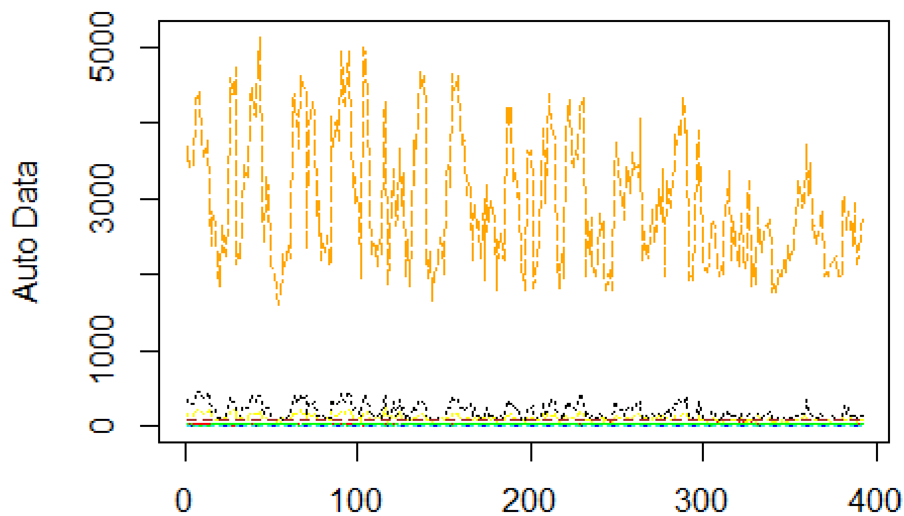
i think that we have three out lievs points and there is no high leverge points. and the QQplots show the normality distrubtion.

End.

4) This question involves the use of multiple linear regression on the Auto data set.

a) Produce a scatterplot matrix which includes all of the variables in the data set.

```
#Exercise forth(ninth of Book):
#a):
matplot(Auto[1:8],col = c("Red","Blue","Black","yellow","orange","Green","Bro
wn",85),ylab = "Auto Data",type = "l")
```



b) Compute the matrix of correlations between the variables using the function cor(). You will need to exclude the name variable, cor() which is qualitative.

```
#b)
(as.matrix(cor(Auto[1:8])))

##                     mpg  cylinders displacement horsepower     weight
## mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442
## cylinders    -0.7776175  1.0000000    0.9508233  0.8429834  0.8975273
## displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944
## horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377
```

```
## weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000
## acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392
## year          0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199
## origin        0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054
##              acceleration       year     origin
## mpg             0.4233285  0.5805410  0.5652088
## cylinders      -0.5046834 -0.3456474 -0.5689316
## displacement   -0.5438005 -0.3698552 -0.6145351
## horsepower     -0.6891955 -0.4163615 -0.4551715
## weight         -0.4168392 -0.3091199 -0.5850054
## acceleration    1.0000000  0.2903161  0.2127458
## year            0.2903161  1.0000000  0.1815277
## origin          0.2127458  0.1815277  1.0000000
```

c) Use the lm() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summary() function to print the results. Comment on the output. For instance:

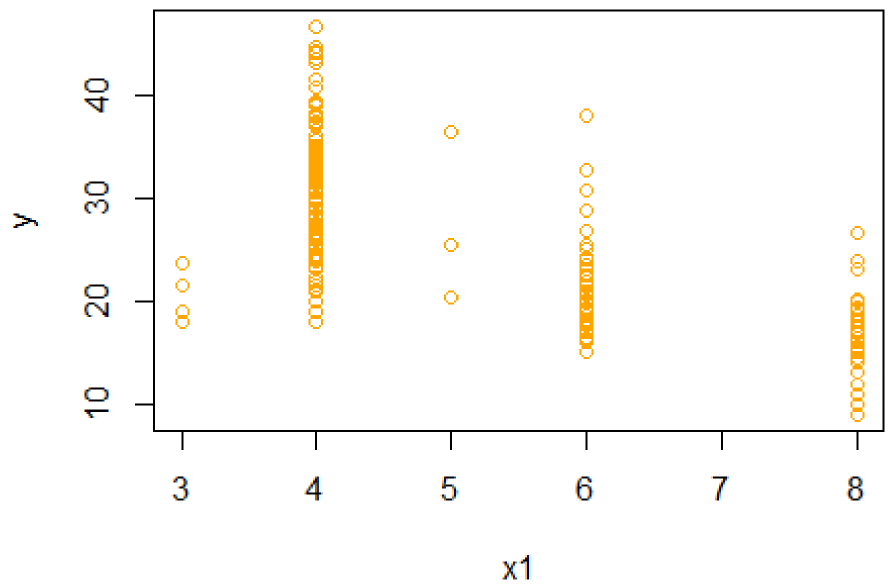
```
#c)
y<-Auto$mpg
x1<-Auto$cylinders
x2<-Auto$displacement
x3<-Auto$horsepower
x4<-Auto$weight
x5<-Auto$acceleration
```

```r
x6<-Auto$year
x7<-Auto$origin
fit1<-lm(y~x1+x2+x3+x4+x5+x6+x7)
summary(fit1)
```

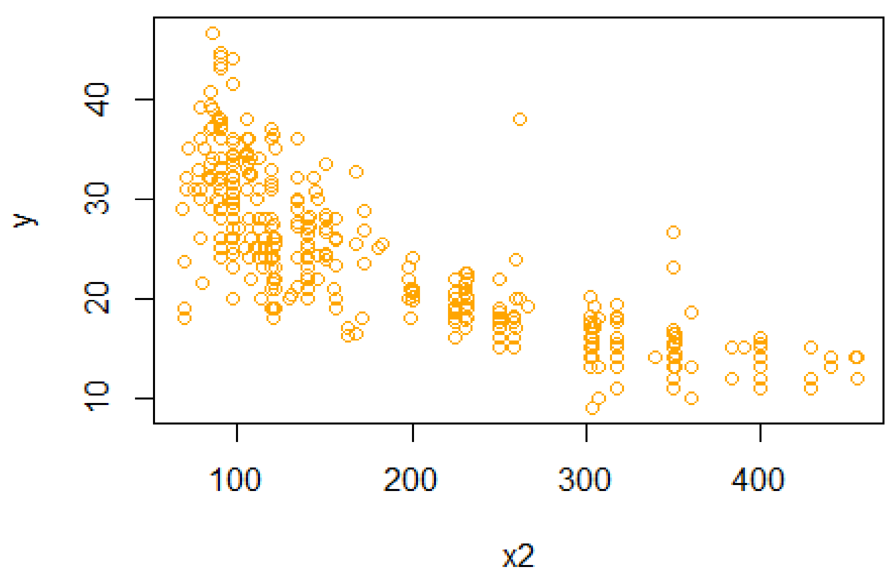
```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5903 -2.1565 -0.1169  1.8690 13.0604
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.218435   4.644294  -3.707  0.00024 ***
## x1           -0.493376   0.323282  -1.526  0.12780
## x2            0.019896   0.007515   2.647  0.00844 **
## x3           -0.016951   0.013787  -1.230  0.21963
## x4           -0.006474   0.000652  -9.929  < 2e-16 ***
## x5            0.080576   0.098845   0.815  0.41548
## x6            0.750773   0.050973  14.729  < 2e-16 ***
## x7            1.426141   0.278136   5.127 4.67e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.328 on 384 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8182
## F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

i. Is there a relationship between the predictors and the response?

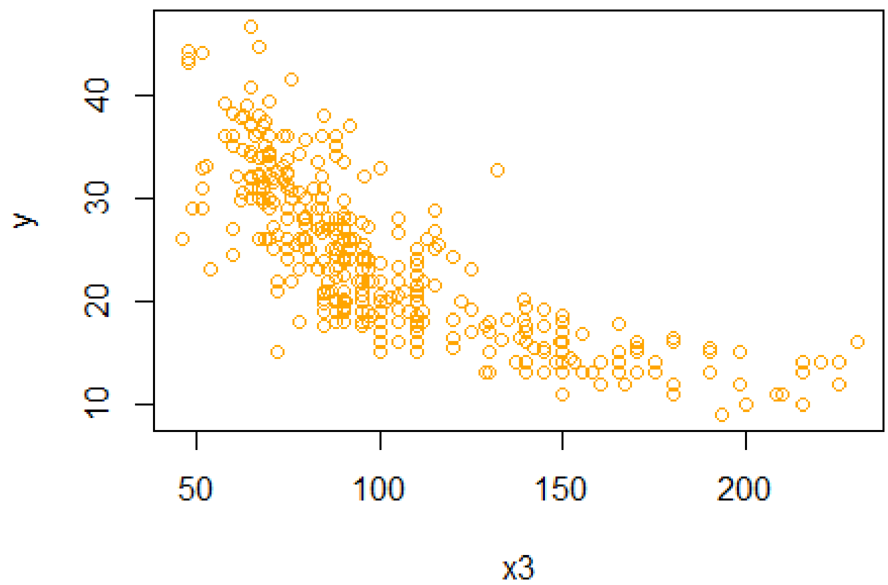
```r
#i)
plot(x1,y,col="orange")
```



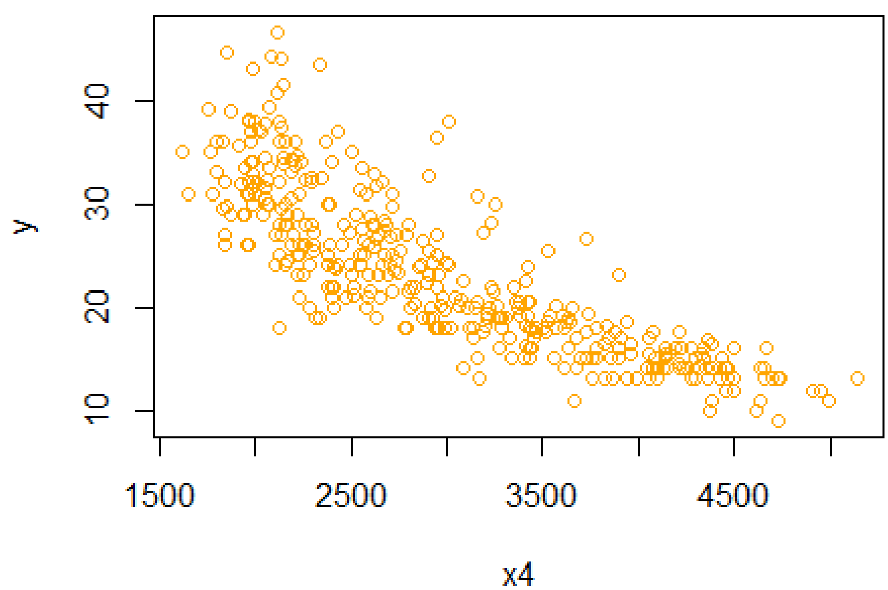

```
plot(x2,y,col="orange")
```



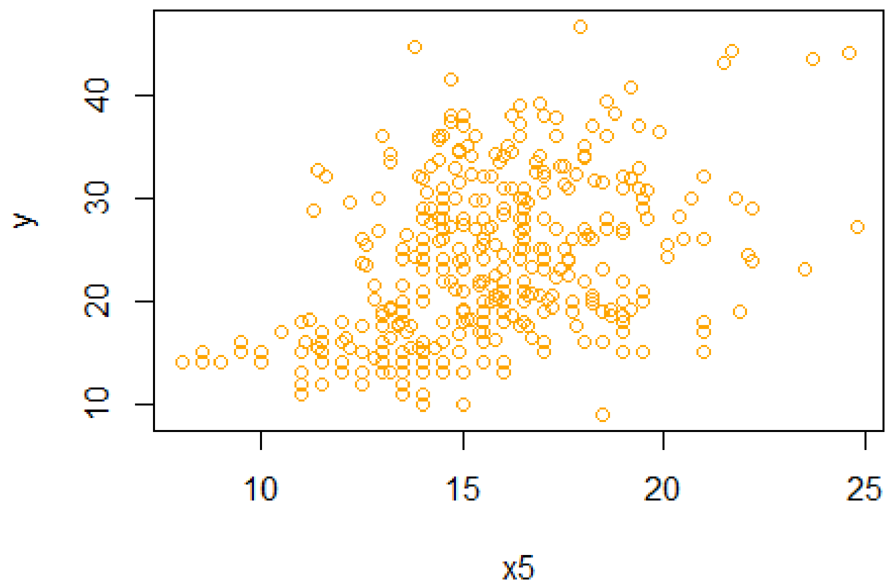

```
plot(x3,y,col="orange")
```



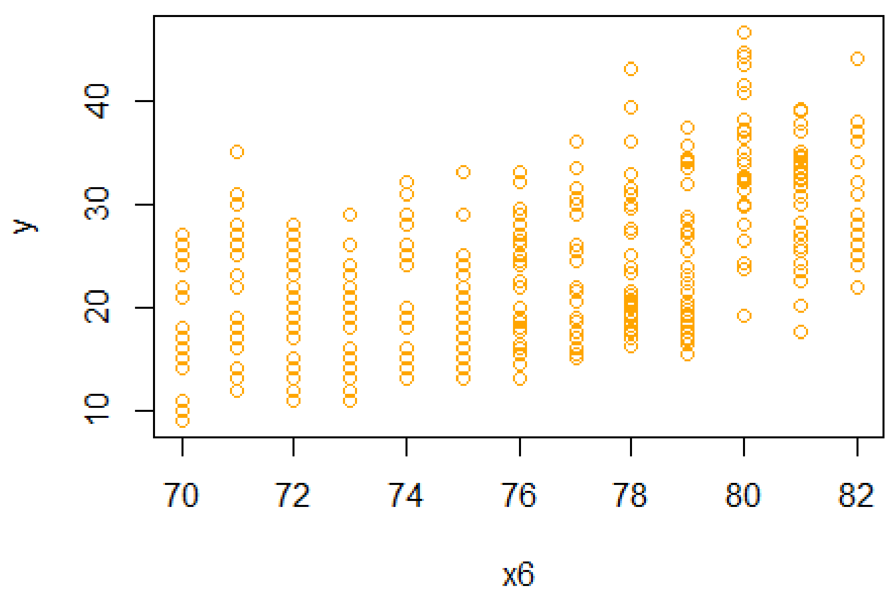

```
plot(x4,y,col="orange")
```



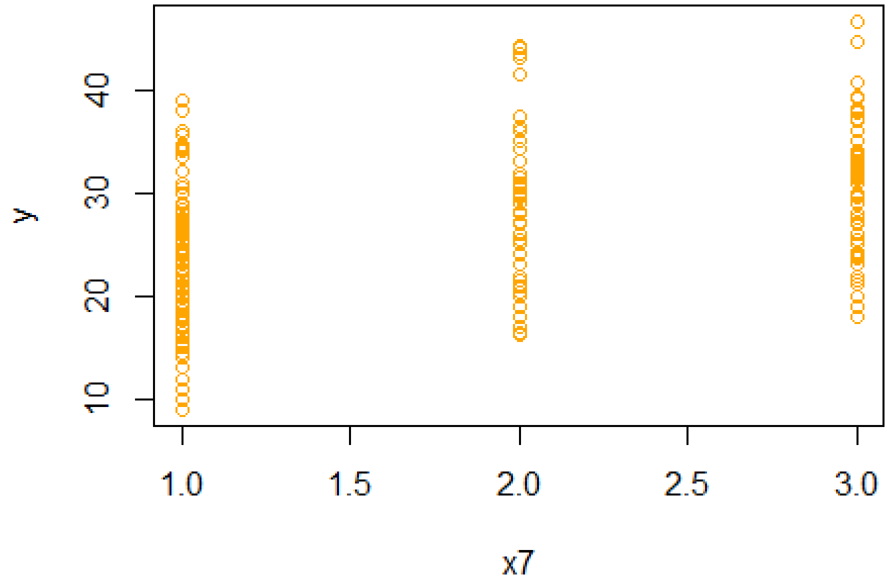

```
plot(x5,y,col="orange")
```




```
plot(x6,y,col="orange")
```




```
plot(x7,y,col="orange")
```




```r
summary(fit1)[8]

## $r.squared
## [1] 0.8214781
```

X2,X3,X4 have negative linear relationship with resonse.

X5 have posetvie relation with response.

ii. Which predictors appear to have a statistically significant relationship to the response?

```r
#ii)
summary(fit1)[4]

## $coefficients
##                 Estimate    Std. Error   t value     Pr(>|t|)
## (Intercept) -17.218434622 4.6442941494 -3.707438 2.401841e-04
## x1           -0.493376319 0.3232823146 -1.526147 1.277965e-01
## x2            0.019895644 0.0075150792  2.647430 8.444649e-03
## x3           -0.016951144 0.0137868914 -1.229512 2.196328e-01
## x4           -0.006474043 0.0006520478 -9.928787 7.874953e-21
## x5            0.080575838 0.0988449567  0.815174 4.154780e-01
## x6            0.750772678 0.0509731223 14.728795 3.055983e-39
## x7            1.426140495 0.2781360924  5.127492 4.665681e-07



anova(fit1)
```

```
## Analysis of Variance Table
##
## Response: y
##            Df  Sum Sq Mean Sq   F value    Pr(>F)
## x1          1 14403.1 14403.1 1300.6838 < 2.2e-16 ***
## x2          1  1073.3  1073.3   96.9293 < 2.2e-16 ***
## x3          1   403.4   403.4   36.4301 3.731e-09 ***
## x4          1   975.7   975.7   88.1137 < 2.2e-16 ***
## x5          1     1.0     1.0    0.0872    0.7679
## x6          1  2419.1  2419.1  218.4609 < 2.2e-16 ***
## x7          1   291.1   291.1   26.2912 4.666e-07 ***
## Residuals 384  4252.2    11.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

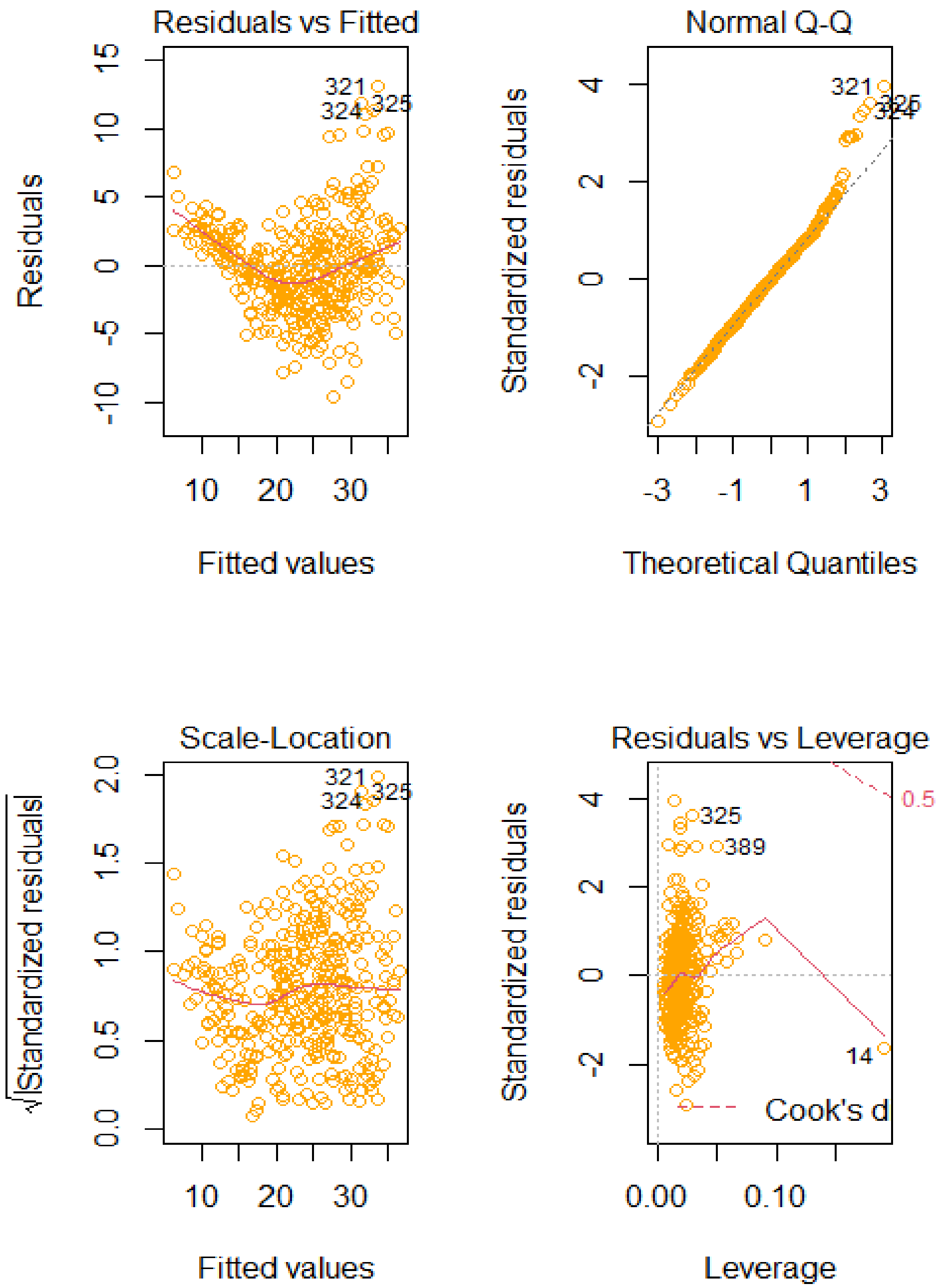
all of the predictors have statistically significant relationship to the response Except X5.

iii. What does the coefficient for the year variable suggest?

```
#iii)
coefficients(fit1)[7]

##        x6
## 0.7507727
```

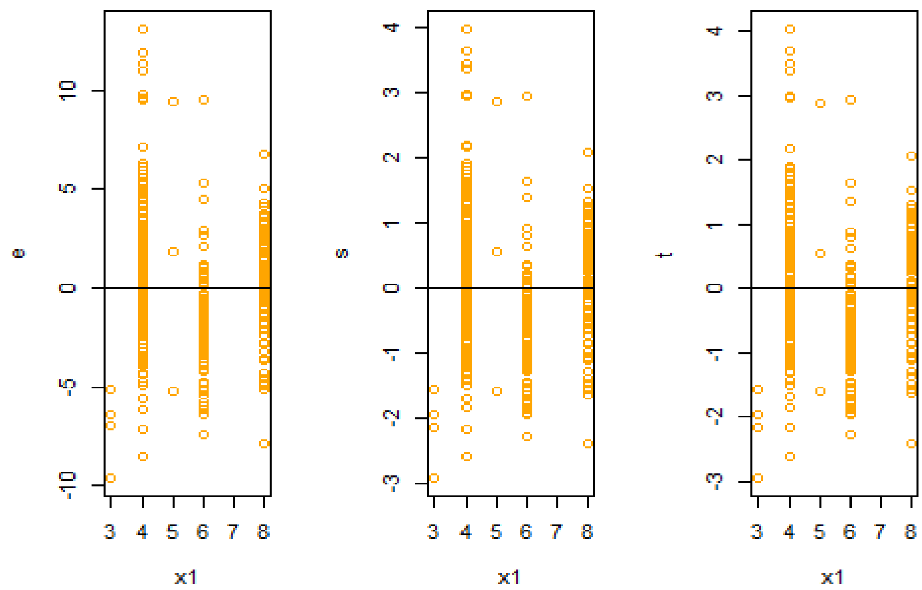
d) Use the plot() function to produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit.Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?

```
#d)
par(mfrow= c(1,2))
plot(fit1,col="Orange")
```

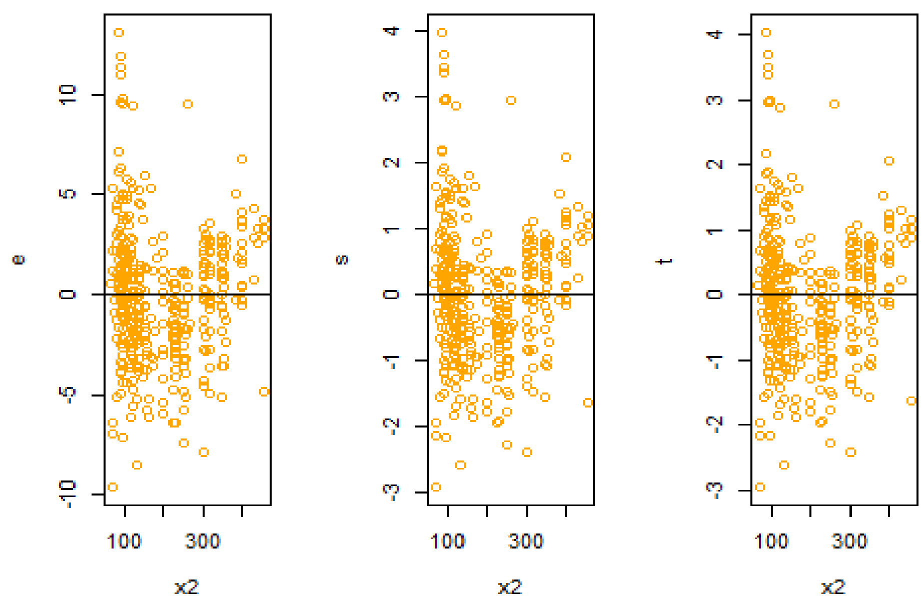



```
e<-residuals(fit1)
s<-rstandard(fit1)
t<-rstudent(fit1)
par(mfrow=c(1,3))
plot(x1,e,col="Orange")
abline(h=0,col="black")
plot(x1,s,col="Orange")
abline(h=0,col="black")
```

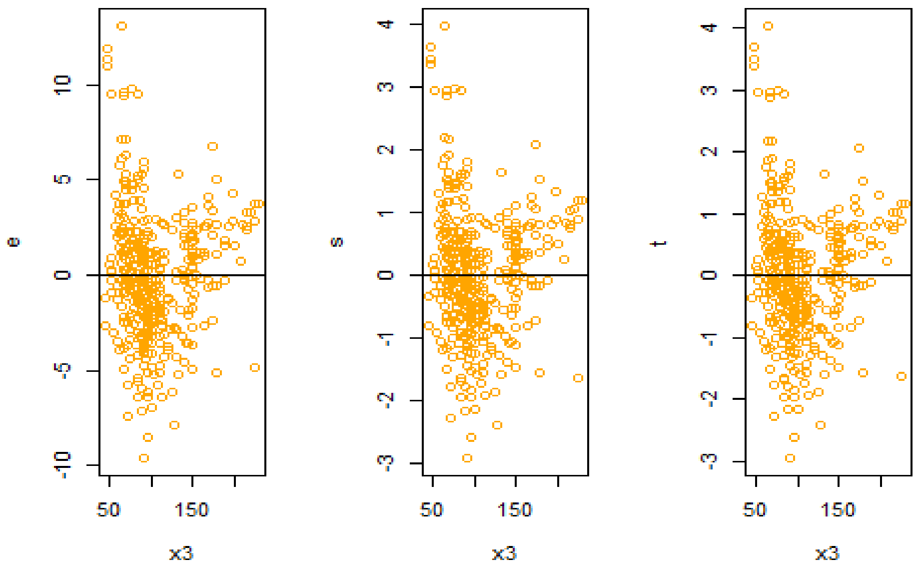
```r
plot(x1,t,col="Orange")
abline(h=0,col="black")
```



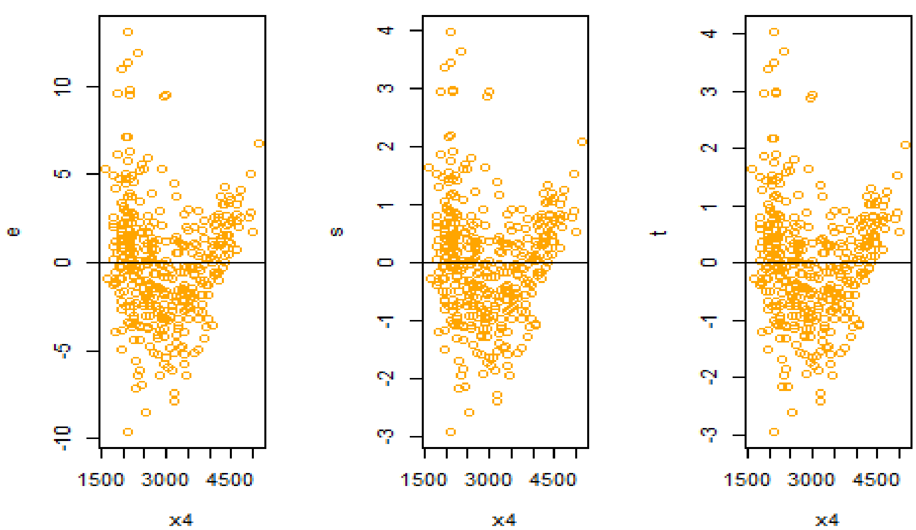

```r
plot(x2,e,col="Orange")
abline(h=0,col="black")
plot(x2,s,col="Orange")
abline(h=0,col="black")
plot(x2,t,col="Orange")
abline(h=0,col="black")
```



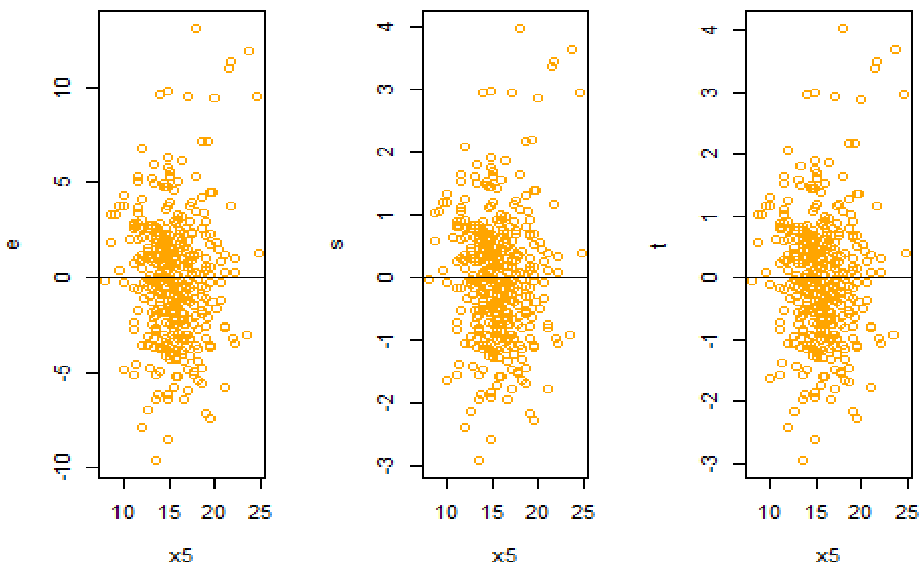
```r
plot(x3,e,col="Orange")
abline(h=0,col="black")
plot(x3,s,col="Orange")
abline(h=0,col="black")
plot(x3,t,col="Orange")
abline(h=0,col="black")
```



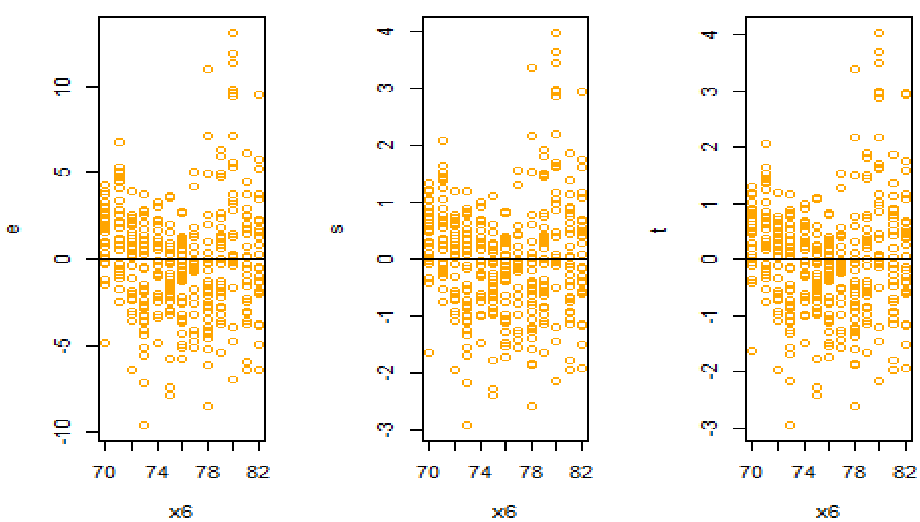
```r
plot(x4,e,col="Orange")
abline(h=0,col="black")
plot(x4,s,col="Orange")
abline(h=0,col="black")
plot(x4,t,col="Orange")
abline(h=0,col="black")
```



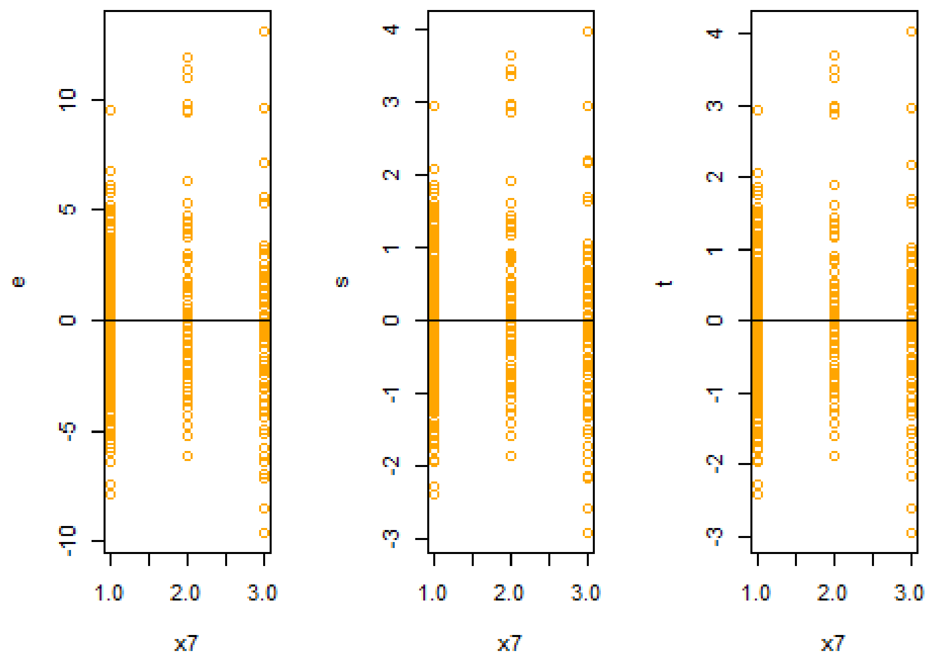
```r
plot(x5,e,col="Orange")
abline(h=0,col="black")
plot(x5,s,col="Orange")
abline(h=0,col="black")
plot(x5,t,col="Orange")
abline(h=0,col="black")
```




```r
plot(x6,e,col="Orange")
abline(h=0,col="black")
plot(x6,s,col="Orange")
abline(h=0,col="black")
plot(x6,t,col="Orange")
abline(h=0,col="black")
```



```
plot(x7,e,col="Orange")
abline(h=0,col="black")
plot(x7,s,col="Orange")
abline(h=0,col="black")
plot(x7,t,col="Orange")
abline(h=0,col="black")
```



at the top of the QQplot we can see some unusually points that are not in y=x linear.

we can see that we dont have fixed standard devation and i think that its posetive function of our predictors.

and we can see some outlievs and High leverage points in our model.

e) Use the * and : symbols to fit linear regression models with interaction effects. Do any interactions appear to be statistically significant?

```
#e)
full.fit<-lm(y~x1+x2+x3+x4+x5+x6+x7+x1*x2+x1*x3+x1*x4+x1*x5+x1*x6+x1*x7+x2*x3
+x2*x4+x3*x5+x3*x6+x3*x7+x4*x5+x4*x6+x4*x7+x5*x6+x5*x7+x6*x7)
summary(full.fit)
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x1 * x2 +
##     x1 * x3 + x1 * x4 + x1 * x5 + x1 * x6 + x1 * x7 + x2 * x3 +
##     x2 * x4 + x3 * x5 + x3 * x6 + x3 * x7 + x4 * x5 + x4 * x6 +
##     x4 * x7 + x5 * x6 + x5 * x7 + x6 * x7)
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5267 -1.4631  0.0026  1.3259 11.3919
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.748e+01  5.028e+01   1.541  0.12414
## x1          -5.346e+00  5.546e+00  -0.964  0.33567
## x2          -1.642e-02  3.542e-02  -0.464  0.64326
## x3           4.370e-01  3.443e-01   1.269  0.20519
## x4          -2.289e-02  1.439e-02  -1.591  0.11251
## x5          -4.954e+00  2.142e+00  -2.313  0.02127 *
## x6           1.802e-01  5.804e-01   0.310  0.75639
## x7          -1.713e+01  6.869e+00  -2.494  0.01308 *
## x1:x2       -4.154e-03  4.452e-03  -0.933  0.35143
## x1:x3        1.165e-02  1.588e-02   0.734  0.46363
## x1:x4        7.368e-04  7.933e-04   0.929  0.35364
## x1:x5        1.346e-01  9.843e-02   1.367  0.17241
## x1:x6       -1.544e-03  6.482e-02  -0.024  0.98101
## x1:x7        6.039e-01  4.263e-01   1.417  0.15747
## x2:x3       -1.977e-04  2.104e-04  -0.940  0.34791
## x2:x4        1.939e-05  1.034e-05   1.876  0.06150 .
## x3:x5       -7.266e-03  3.641e-03  -1.996  0.04669 *
## x3:x6       -5.333e-03  3.949e-03  -1.350  0.17772
## x3:x7       -2.921e-03  2.914e-02  -0.100  0.92020
## x4:x5        1.719e-04  1.801e-04   0.954  0.34051
## x4:x6        7.777e-05  1.776e-04   0.438  0.66165
## x4:x7        8.030e-04  1.146e-03   0.701  0.48376
## x5:x6        4.802e-02  2.521e-02   1.905  0.05754 .
## x5:x7        4.607e-01  1.449e-01   3.179  0.00161 **
## x6:x7        7.784e-02  7.101e-02   1.096  0.27372
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.714 on 367 degrees of freedom
## Multiple R-squared:  0.8865, Adjusted R-squared:  0.8791
## F-statistic: 119.4 on 24 and 367 DF,  p-value: < 2.2e-16
```

```
anova(full.fit)

## Analysis of Variance Table
##
## Response: y
##            Df  Sum Sq Mean Sq   F value    Pr(>F)
## x1          1 14403.1 14403.1 1955.3643 < 2.2e-16 ***
## x2          1  1073.3  1073.3  145.7173 < 2.2e-16 ***
## x3          1   403.4   403.4   54.7667 9.343e-13 ***
## x4          1   975.7   975.7  132.4645 < 2.2e-16 ***
## x5          1     1.0     1.0    0.1312 0.7174471
## x6          1  2419.1  2419.1  328.4201 < 2.2e-16 ***
## x7          1   291.1   291.1   39.5245 9.203e-10 ***
## x1:x2       1   596.8   596.8   81.0189 < 2.2e-16 ***
## x1:x3       1   370.8   370.8   50.3382 6.715e-12 ***
## x1:x4       1    36.0    36.0    4.8862 0.0276888 *
## x1:x5       1     4.9     4.9    0.6609 0.4167787
## x1:x6       1    97.9    97.9   13.2872 0.0003058 ***
## x1:x7       1    42.5    42.5    5.7724 0.0167756 *
## x2:x3       1    76.8    76.8   10.4282 0.0013530 **
## x2:x4       1     5.1     5.1    0.6931 0.4056486
## x3:x5       1    44.7    44.7    6.0664 0.0142361 *
## x3:x6       1    90.8    90.8   12.3300 0.0005013 ***
## x3:x7       1    33.3    33.3    4.5227 0.0341142 *
## x4:x5       1     4.2     4.2    0.5665 0.4521240
## x4:x6       1     7.7     7.7    1.0495 0.3062885
## x4:x7       1    45.2    45.2    6.1410 0.0136574 *
## x5:x6       1    17.7    17.7    2.4024 0.1220083
## x5:x7       1    65.7    65.7    8.9155 0.0030173 **
## x6:x7       1     8.9     8.9    1.2016 0.2737230
## Residuals 367  2703.3     7.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

now we can see that or p_value for predictors are signifact when is lower than 0.05.

the bethaj for j=0,1,2,3,4,5,6,7 is equal to the Pr(>|t|) column.

the Residual standard error = 2.714, R-squared = 0.8791.

f) Try a few different transformations of the variables, such as log(X), √X, X2. Comment on your findings.

```
#f)
#the log(X) transformation:
y<-Auto$mpg
x1<-log(Auto$cylinders)
x2<-log(Auto$displacement)
x3<-log(Auto$horsepower)
x4<-log(Auto$weight)
x5<-log(Auto$acceleration)
x6<-log(Auto$year)
x7<-log(Auto$origin)
fit4<-lm(y~x1+x2+x3+x4+x5+x6+x7)
summary(fit4)

##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5987 -1.8172 -0.0181  1.5906 12.8132
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -66.5643    17.5053  -3.803 0.000167 ***
## x1            1.4818     1.6589   0.893 0.372273
## x2           -1.0551     1.5385  -0.686 0.493230
## x3           -6.9657     1.5569  -4.474 1.01e-05 ***
## x4          -12.5728     2.2251  -5.650 3.12e-08 ***
## x5           -4.9831     1.6078  -3.099 0.002082 **
## x6           54.9857     3.5555  15.465  < 2e-16 ***
## x7            1.5822     0.5083   3.113 0.001991 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.069 on 384 degrees of freedom
## Multiple R-squared:  0.8482, Adjusted R-squared:  0.8454
## F-statistic: 306.5 on 7 and 384 DF,  p-value: < 2.2e-16
```

now we can see that or p_value for predictors are signifact when is lower than 0.05.

the bethaj for j=0,1,2,3,4,5,6,7 is equal to the Pr(>|t|) column.

the Residual standard error =3.069 , R-squared =0.8482 .

```r
#the second root of X transformation:
y<-Auto$mpg
x1<-sqrt(Auto$cylinders)
x2<-sqrt(Auto$displacement)
x3<-sqrt(Auto$horsepower)
x4<-sqrt(Auto$weight)
x5<-sqrt(Auto$acceleration)
x6<-sqrt(Auto$year)
x7<-sqrt(Auto$origin)
fit5<-lm(y~x1+x2+x3+x4+x5+x6+x7)
summary(fit5)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5250 -1.9822 -0.1111  1.7347 13.0681
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -49.79814    9.17832  -5.426 1.02e-07 ***
## x1           -0.23699    1.53753  -0.154   0.8776
## x2            0.22580    0.22940   0.984   0.3256
## x3           -0.77976    0.30788  -2.533   0.0117 *
## x4           -0.62172    0.07898  -7.872 3.59e-14 ***
## x5           -0.82529    0.83443  -0.989   0.3233
## x6           12.79030    0.85891  14.891  < 2e-16 ***
## x7            3.26036    0.76767   4.247 2.72e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.21 on 384 degrees of freedom
## Multiple R-squared:  0.8338, Adjusted R-squared:  0.8308
## F-statistic: 275.3 on 7 and 384 DF,  p-value: < 2.2e-16
```

now we can see that or p_value for predictors are signifact when is lower than 0.05.

the bethaj for j=0,1,2,3,4,5,6,7 is equal to the Pr(>|t|) column.

the Residual standard error = 3.21, R-squared = 0.8338.

```r
#the X power to two transformation:
y<-Auto$mpg
x1<-(Auto$cylinders)^2
x2<-(Auto$displacement)^2
x3<-(Auto$horsepower)^2
x4<-(Auto$weight)^2
x5<-(Auto$acceleration)^2
x6<-(Auto$year)^2
x7<-(Auto$origin)^2
fit6<-lm(y~x1+x2+x3+x4+x5+x6+x7)
summary(fit6)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6786 -2.3227 -0.0582  1.9073 12.9807
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.208e+00  2.356e+00   0.513 0.608382
## x1          -8.829e-02  2.521e-02  -3.502 0.000515 ***
## x2           5.680e-05  1.382e-05   4.109 4.87e-05 ***
## x3          -3.621e-05  4.975e-05  -0.728 0.467201
## x4          -9.351e-07  8.978e-08 -10.416  < 2e-16 ***
## x5           6.278e-03  2.690e-03   2.334 0.020130 *
## x6           4.999e-03  3.530e-04  14.160  < 2e-16 ***
## x7           4.129e-01  6.914e-02   5.971 5.37e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.539 on 384 degrees of freedom
## Multiple R-squared:  0.7981, Adjusted R-squared:  0.7944
## F-statistic: 216.8 on 7 and 384 DF,  p-value: < 2.2e-16
```

now we can see that or p_value for predictors are signifact when is lower than 0.05.

the bethaj for j=0,1,2,3,4,5,6,7 is equal to the Pr(>|t|) column.

the Residual standard error = 3.539, R-squared = 0.7981.

End.

5) This question should be answered using the Carseats data set.

a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

```
#Exercise fivth(Ten of Book):
#introducing the variables:for classifactions qualitative variables(Yes=2
, No=1)

#a)
y<-Carseats$Sales
x1<-Carseats$Price
x2<-as.integer(Carseats$Urban)
x3<-as.integer(Carseats$US)
fit1<-lm(y~x1+x2+x3)
```

yes =2 and no =1

b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!

```
#b)
summary(fit1)

##
## Call:
## lm(formula = y ~ x1 + x2 + x3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9206 -1.6220 -0.0564  1.5786  7.0581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.864812   0.841681  14.097  < 2e-16 ***
## x1          -0.054459   0.005242 -10.389  < 2e-16 ***
## x2          -0.021916   0.271650  -0.081    0.936
## x3           1.200573   0.259042   4.635 4.86e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

just X3 have posetive relationship with response.

c) Write out the model in equation form, being careful to handle the qualitative variables properly.

```
#c)
f<-function(X1,X2,X3){
  f<-(as.numeric(coef(fit1)[1])+ (as.numeric(coef(fit1)[2])*X1)+ (as.numeric(
coef(fit1)[3])*X2)+ (as.numeric(coef(fit1)[4])*X3))
  return(f)
}
```

d) For which of the predictors can you reject the null hypothesis H0 : βj = 0?

```
#d)
anova(fit1)

## Analysis of Variance Table
##
## Response: y
##            Df  Sum Sq Mean Sq  F value    Pr(>F)
## x1          1  630.03  630.03 103.0603 < 2.2e-16 ***
## x2          1    0.10    0.10   0.0158    0.9001
## x3          1  131.31  131.31  21.4802  4.86e-06 ***
## Residuals 396 2420.83    6.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit1)[5]

##           Pr(>F)
## x1        <2e-16 ***
## x2        0.9001
## x3        <2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

now we can see that all of our predictors are signifact Except X2.

e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

```
#e)
Reduce.fit<-lm(Sales~US+Price,data = Carseats)
```

f) How well do the models in (a) and (e) fit the data?

```
#f)
summary(Reduce.fit)
```

```
##
## Call:
## lm(formula = Sales ~ US + Price, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9269 -1.6286 -0.0574  1.5766  7.0515
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.03079    0.63098  20.652  < 2e-16 ***
## USYes        1.19964    0.25846   4.641 4.71e-06 ***
## Price       -0.05448    0.00523 -10.416  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.469 on 397 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2354
## F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

the R squared is fixed(0.2393) when we remove the Urban perdictors form our model.

the residuals standard error just is lower(0.003) than the full model.

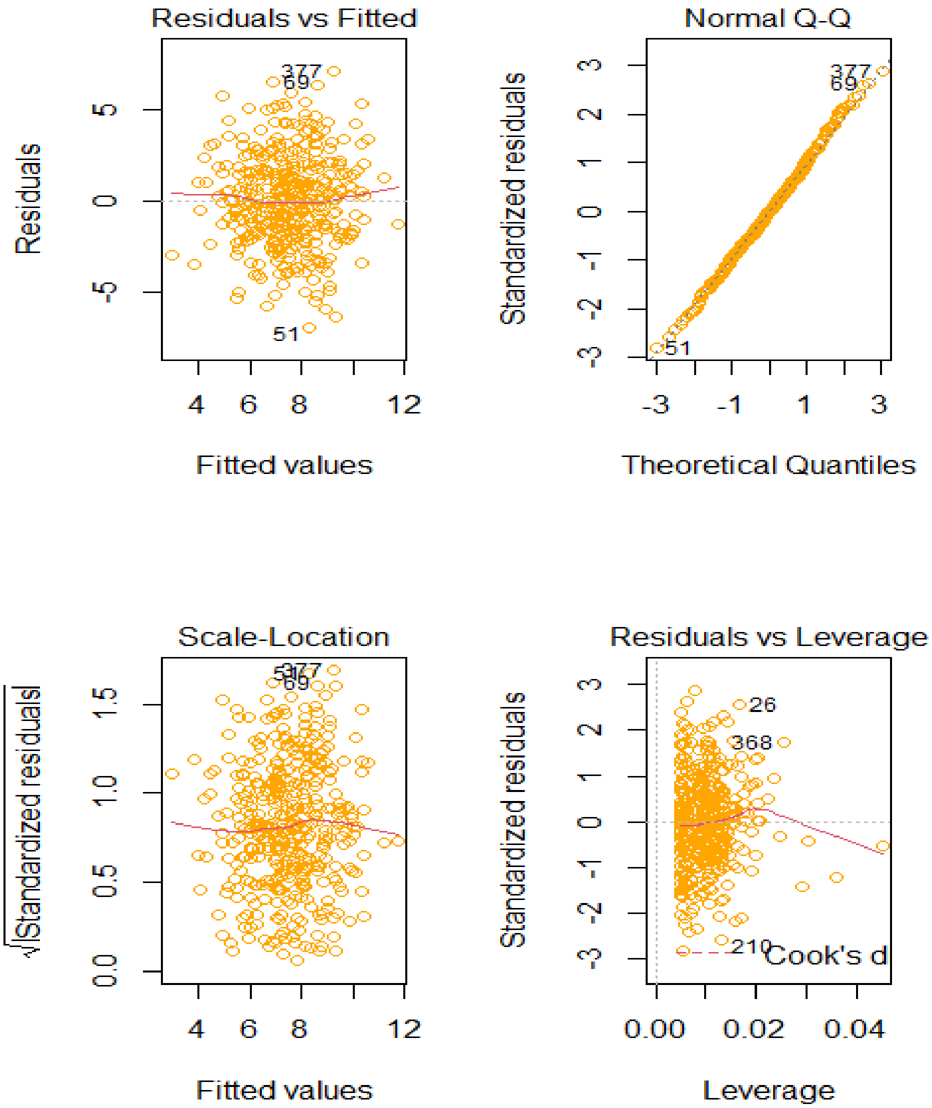
g) Using the model from (e), obtain 95 % confidence intervals for the coefficient(s).

```
#g)
confint(fit1,level = 0.95)
```

```
##                   2.5 %      97.5 %
## (Intercept) 10.21009150 13.51953328
## x1          -0.06476419 -0.04415351
## x2          -0.55597316  0.51214085
## x3           0.69130419  1.70984121
```

h) Is there evidence of outliers or high leverage observations in the model from (e)?

```
#h)
par(mfrow=c(1,2))
plot(fit1)
```



i think that our residuals have standard normal distrubtion.

at the first devation is low then more after that again lower.

we have some lievs andHigh leverage points.

End.