組別：12

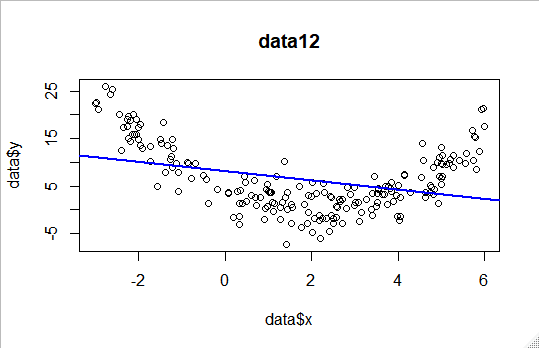
組員：b07801003梁嫚芳

b07801022陳沛宜

b07801054劉昀真

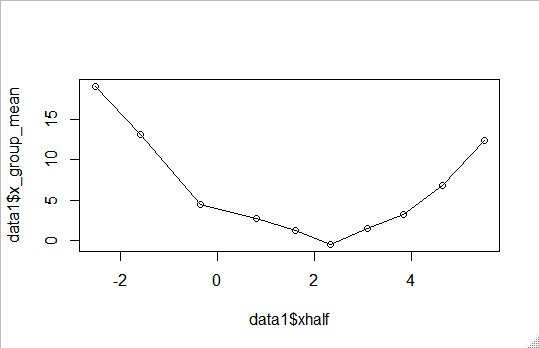
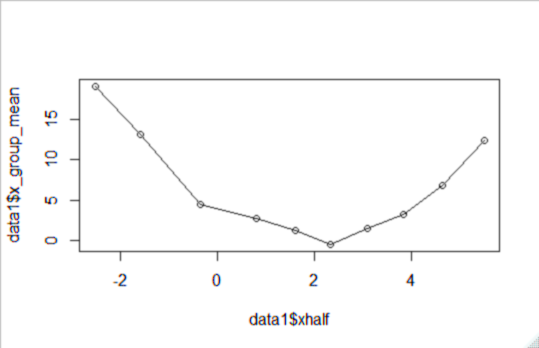
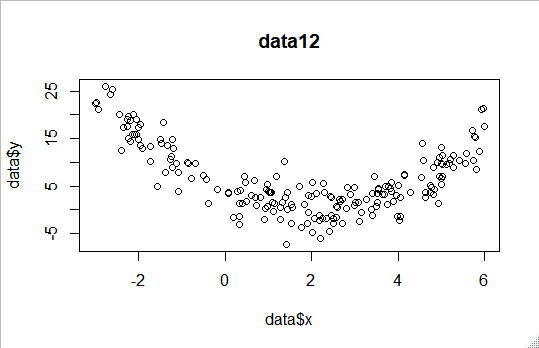
1.

(1) y=-0.9781x +8.1803



(2) 此模型無法有效解釋變數x與y的關係。首先，報表中的Adjusted R-squared=0.1234遠小於1，另可由散佈圖的分布看出較偏向拋物線，而非線性模型中的一次函數圖形。且報表中Residual standard error = 6.497遠高於0，代表MSE^2也遠高於0，意味模型對於解釋x與y的關係並無幫助。

2.

(1)

(2)

3.

(1)

problem-1求出模型中的y值=**10.1365**

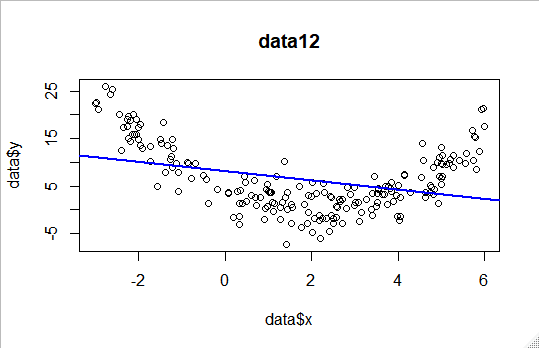
Problem-2報表：

x\_group\_mean xhalf

1 19.1474500 -2.521500

2 13.1731800 -1.597450

利用此報表兩點以內插法求得當x(xhalf)=-2時的預測值y(x\_group\_mean)=**15.77579066**

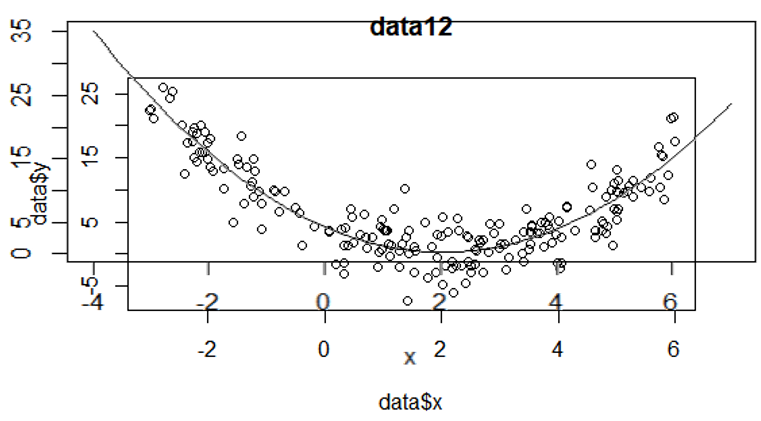
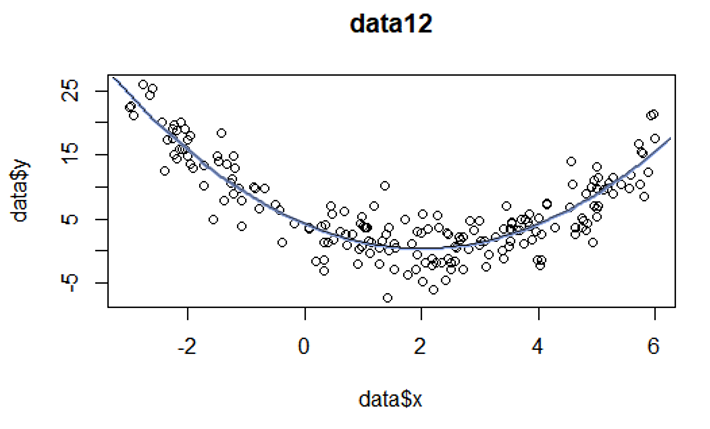
(2) Problem-2的方式較佳。

(3) 因為由第一題的模型與散布圖中看出，當x=-2時資料點幾乎分布於迴歸線上方，若用problem-1的方式求出的預測值則與資料群落差較大，而Problem-2分段方式所求之預測值(15.78)較符合x-y散佈圖上數據分布。

4.

(1) 二次曲線模型

(2) y=4.27659-3.91227\*x+0.95367\*(x^2)

經疊圖後

When x=-2, **y=15.916**

(3) 是。因為此模型的報表顯示Adjusted R-squared=0.8082。

5.

是

令二次迴歸線為y= 𝛽0- 𝛽1\*x+ 𝛽2\*(x^2)

H0: 𝛽1= 𝛽2=0

H1: 𝛽1, 𝛽2不均為0

設𝛼=0.01

而P-value: < 2.2e-16 遠小於0.01

F-statistic: 420.2 on 2 and 197 DF, p-value: < 2.2e-16

6.

Problem-4。

語法：

##1

data<-read.csv(file.choose())

View(data)

model<-lm(data$y~data$x,data)

summary(model)

plot(data$x,data$y,main="data12")

abline(model,col="blue",lwd=2)

##2

cut\_x<-seq(1,200,20)

datax<-data[order(data$x),]

x\_1<-datax[1:20,]

x\_2<-datax[21:40,]

x\_3<-datax[41:60,]

x\_4<-datax[61:80,]

x\_5<-datax[81:100,]

x\_6<-datax[101:120,]

x\_7<-datax[121:140,]

x\_8<-datax[141:160,]

x\_9<-datax[161:180,]

x\_10<-datax[181:200,]

for(i in cut\_x){

half[i/20+19/20]<-(datax$x[i]+datax$x[i+19])/2

}

data1<-data.frame(x\_group\_mean=c(mean(x\_1$y),mean(x\_2$y),mean(x\_3$y),mean(x\_4$y),mean(x\_5$y),mean(x\_6$y),mean(x\_7$y),mean(x\_8$y),mean(x\_9$y),mean(x\_10$y)),xhalf=half[1:10])

plot(data1$xhalf,data1$x\_group\_mean,type="o")

data1

##4

model\_diy<-lm(data$y~data$x+I(data$x^2),data)

summary(model\_diy)

plot(data$x,data$y,main="data12")

curve(4.27659-3.91227\*x+0.95367\*(x^2))

##5

summary(model\_diy)