4. (a)(b)RSE(,)> RSE()，又 所以 train( cubic) RSS < train(linear) RSS ，test( cubic) RSS >test (linear) RSS但可以使用 partial F 來測量是否有相關

假設檢定為

(c)(d),p=2,3,…..，則可知非線性相關，結果應該會跟原本的X不相關(corr很小) 可以適用goodness of fit test 檢測是否有非線性相關，結論train( cubic) RSS < train(linear) RSS，但test( cubic) RSS 、test (linear) RSS 說不準 是誰比較高，因為 p

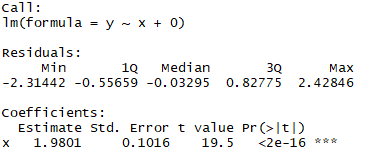
5.,

6. 因為 假設為 且 E()=0 ,

= = 0=0=0

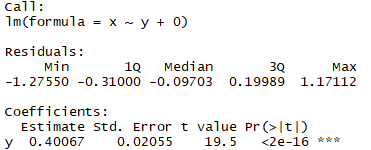
11.

(i)



P<0.05 拒絕

(ii)



P<0.05 拒絕

(iii)

因為原式是 則挪項可得x=

所以估計式殘差會差一個負號 答案會有 1-1的關係

(iv)

,=

其中

=

則可得

(v)

可知 & x= 有一樣的

(vi)由(i)(ii) 可知t-statistic 一樣

**13**

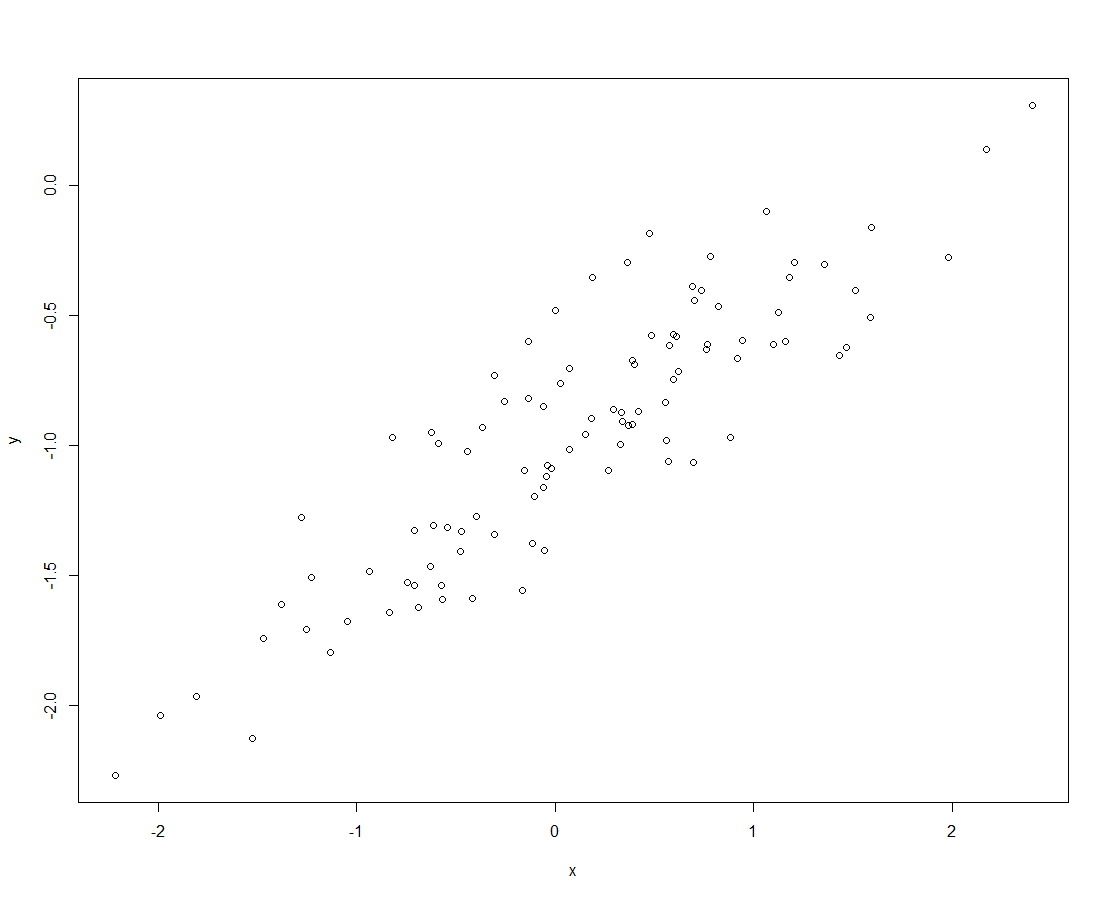
(a)(b)(c)





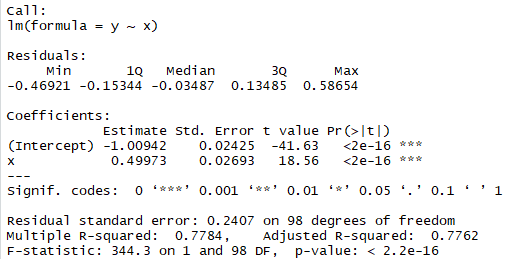
Y有100筆，然而 ,

(d)



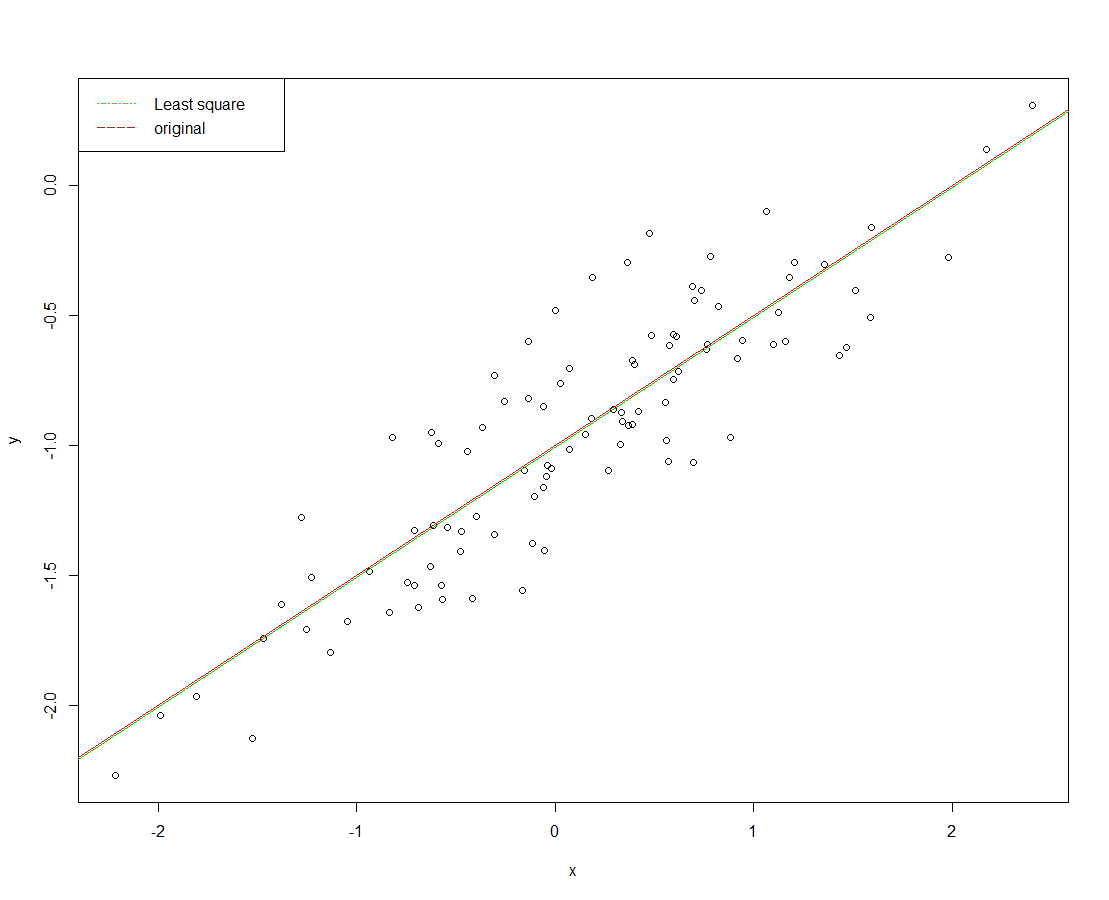
大概可以看出正相關

(e)

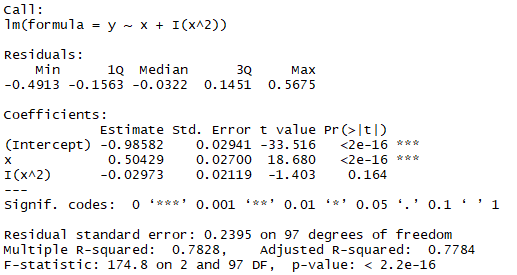


跟原始的設定已經非常接近

(f)



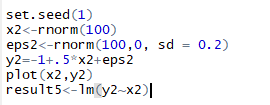
(g)

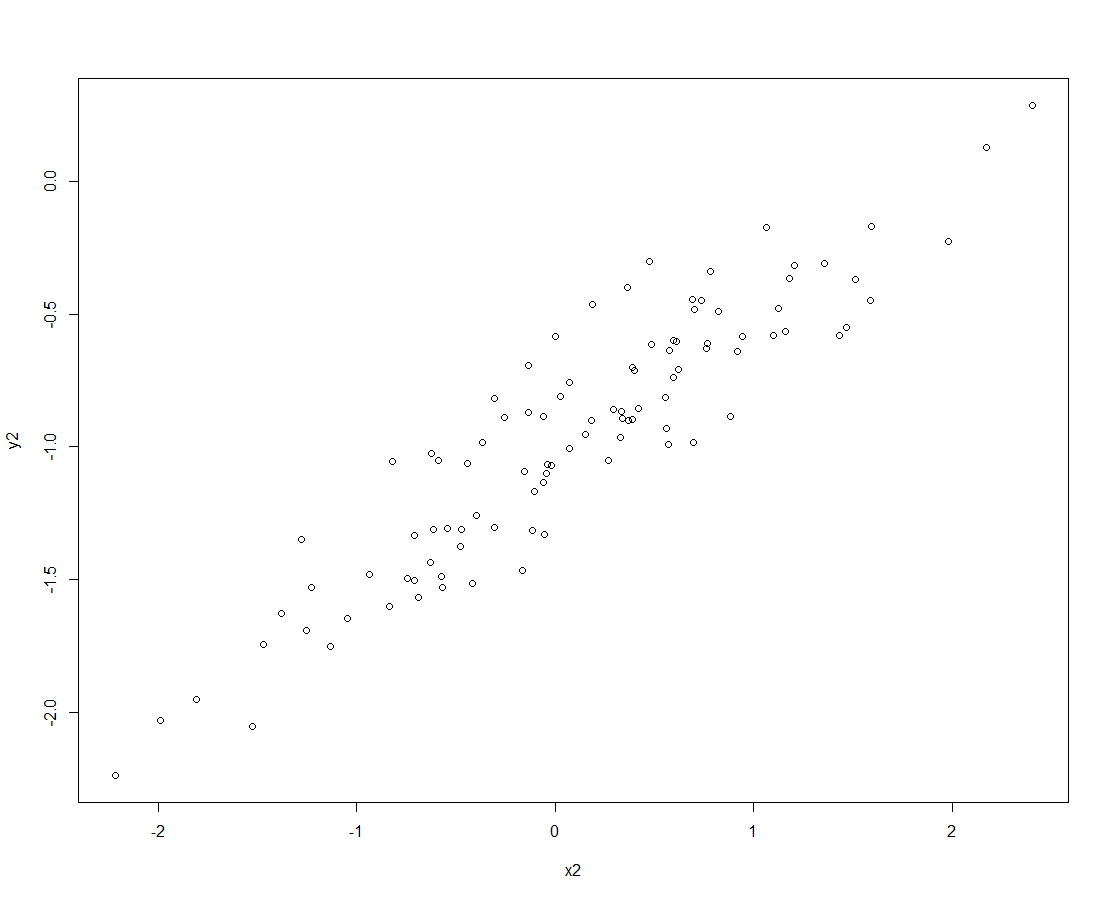


發現 並沒有猛烈變動 因為 p

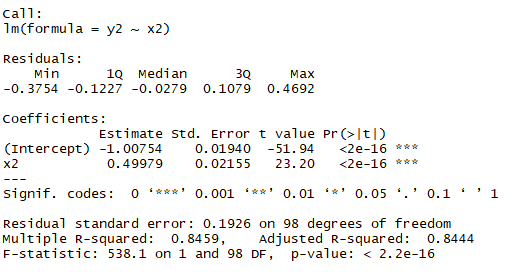
的確增加了 但沒有增加太多 且 並沒有符合假設檢定

(h)



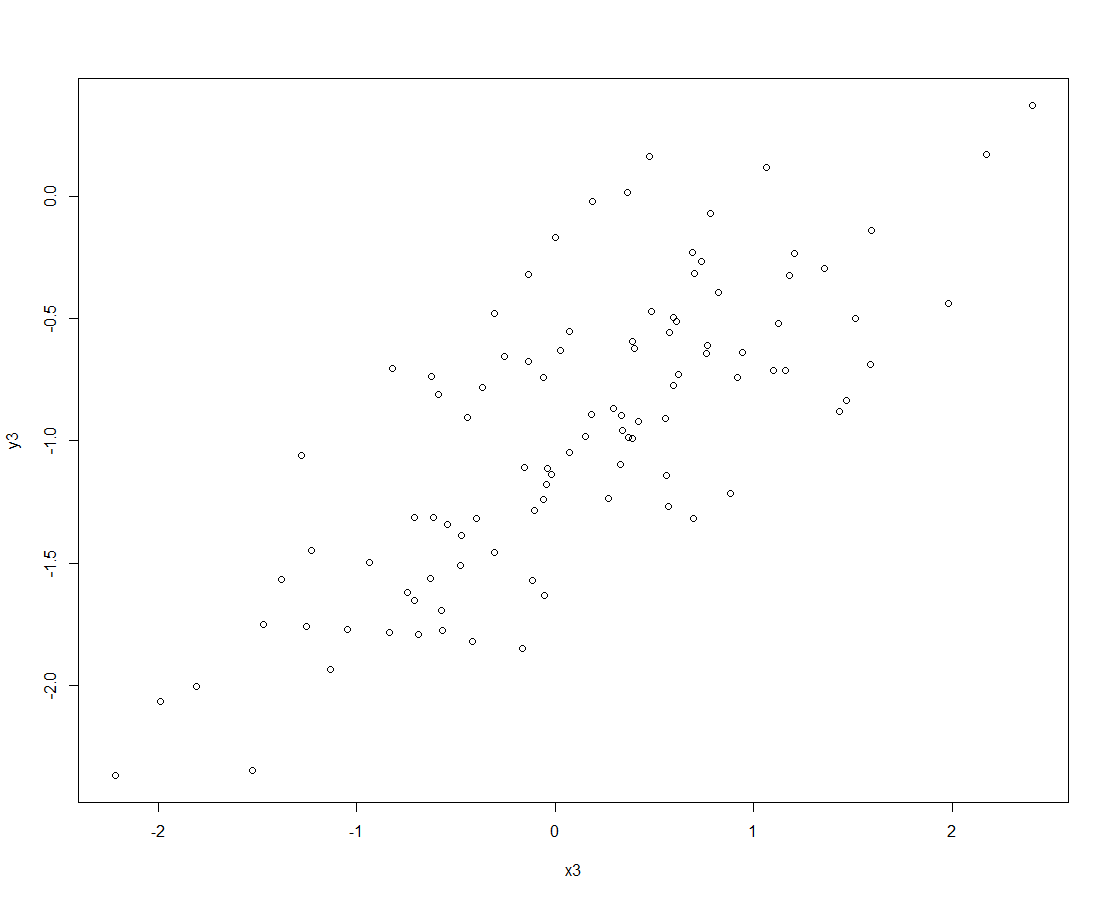


跟上一張圖有更像有較強線性關係

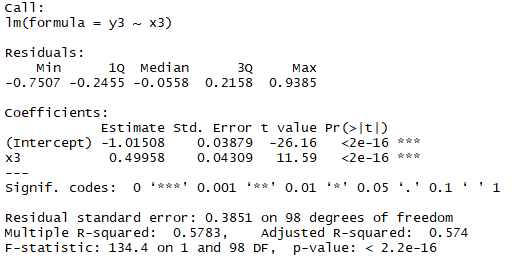


差距不大，但更小了(因為sse 受noise 影響)

(i)

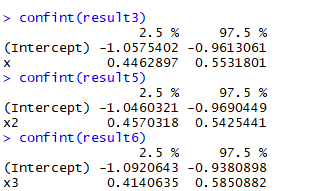


跟上一張圖有較弱線性關係



差距不大，但更大了(因為sse 受noise 影響)

(j)



受noise 跟n的影響(可無限逼近真實)，但 有其上限，這上限跟noise 有關；在noise是一個homogeneous 的狀態，我們可以說我們最好的Model，存在一個上限的，超過的都是overfiting

程式碼

set.seed(1)

x<-rnorm(100)

eps<-rnorm(100,0, sd = 0.25)

y=-1+.5\*x+eps

plot(x,y)

result3<-lm(y~x)

summary(result3)

abline(result3,ltd=4,col=3)

abline(a=-1,b=.5,ltd=5,col=2)

legend("topleft", c("Least square", "original"), col = c("green", "red"), lty = c(4, 5))

result4<-lm(y~x+I(x^2))

summary(result4)

set.seed(1)

x2<-rnorm(100)

eps2<-rnorm(100,0, sd = 0.2)

y2=-1+.5\*x2+eps2

plot(x2,y2)

result5<-lm(y2~x2)

abline(result5,ltd=4,col=3)

abline(a=-1,b=.5,ltd=5,col=2)

legend("topleft", c("Least square", "original"), col = c("green", "red"), lty = c(4, 5))

summary(result5)

set.seed(1)

x3<-rnorm(100)

eps3<-rnorm(100,0, sd = 0.4)

y3=-1+.5\*x3+eps3

plot(x3,y3)

result6<-lm(y3~x3)

summary(result6)

confint(result3)

confint(result5)

confint(result6)