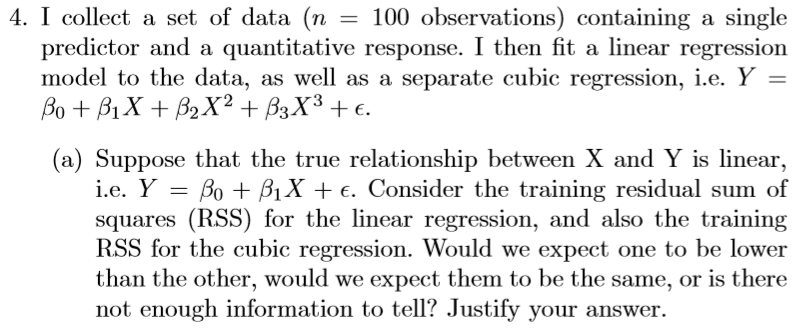
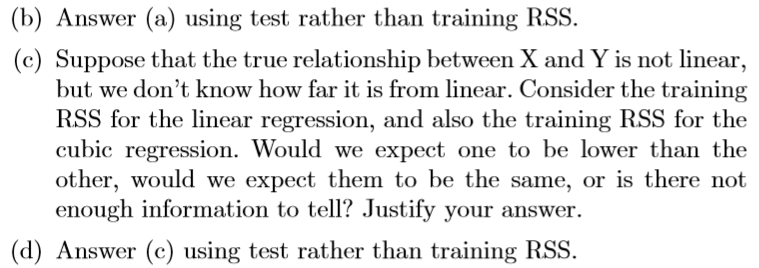
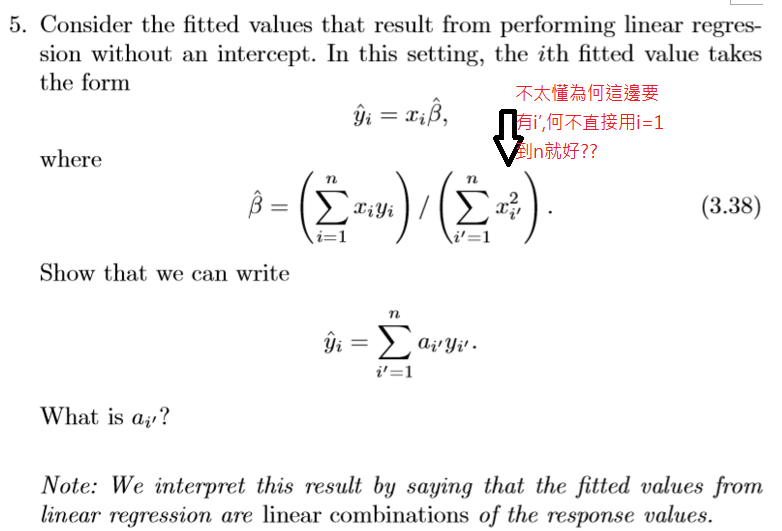
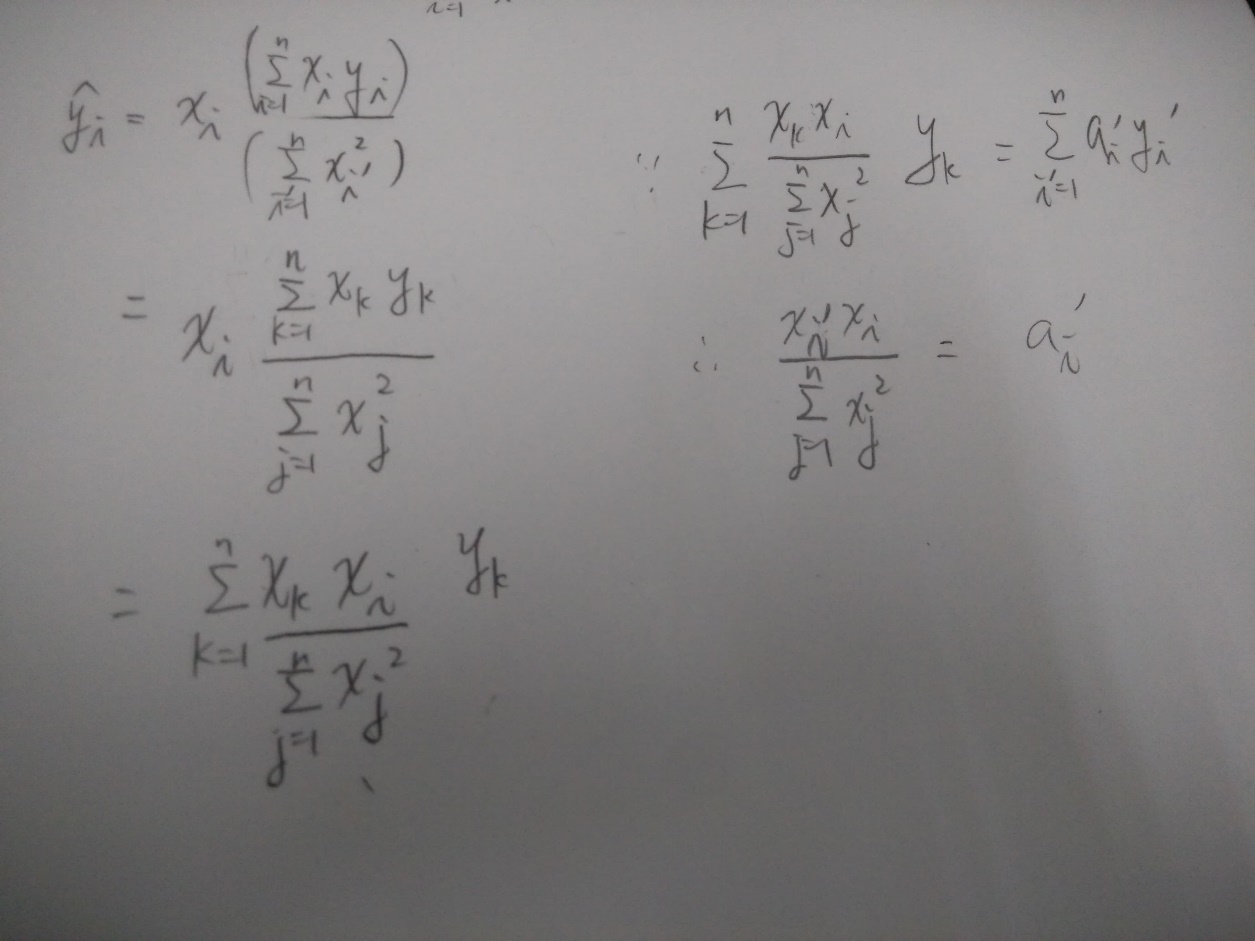
應數碩一 M072040019 梅瀚中

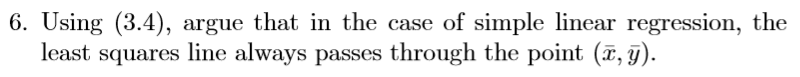


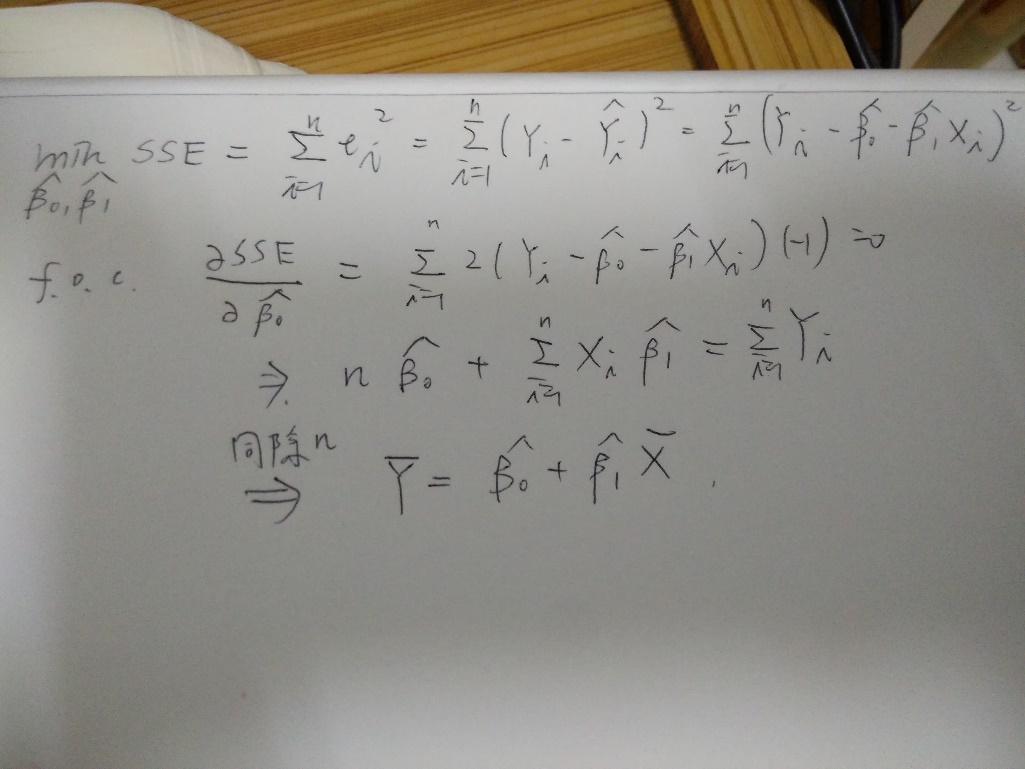


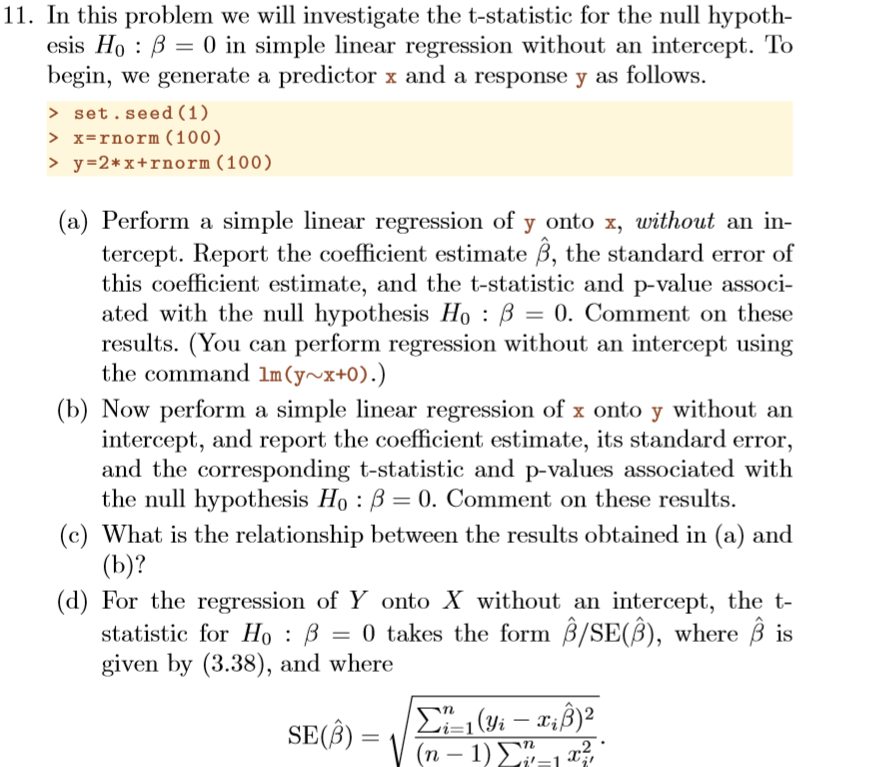
1. 如果真實X和Y的關係是線性的,那使用linear regression在training set的RSS會低於使用cubic regression的RSS。
2. 高次方的regression通常在training set上的準確率通常會很好,但是在testing set上的表現不會太好,稱為overfitting。我認為高次方的testing RSS會比training RSS 更高。
3. 真實X和Y的關係是非線性,那使用cubic regression在training set的RSS會低於使用linear regression的RSS
4. 如果真實X和Y的關係是非線性的,那使用linear regression通常會在training set和testing set上underfitting 。但是我們也不知道它指的”非線性”的程度到底是多大, 如果它更接近線性而不是立方，則linear的testing RSS可能低於cubic的testing RSS。

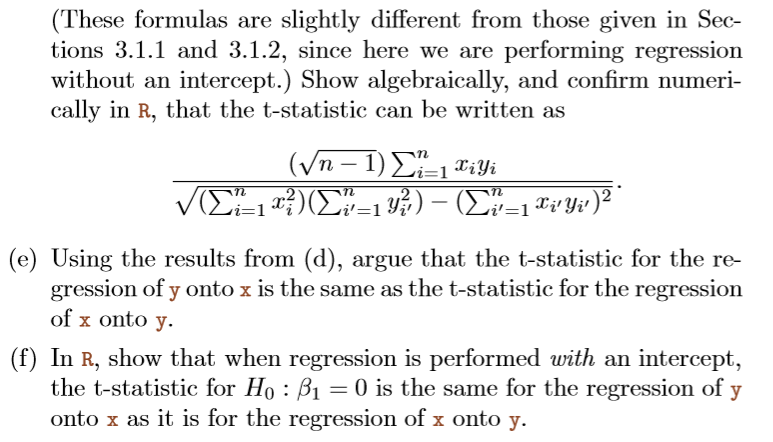




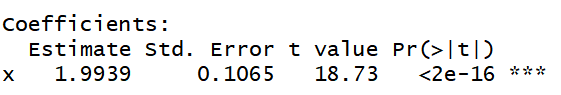






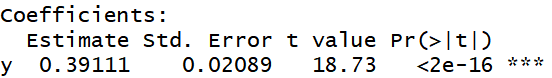


(a)



根據p-value值非常的小,我們有顯著證據拒絕H0。

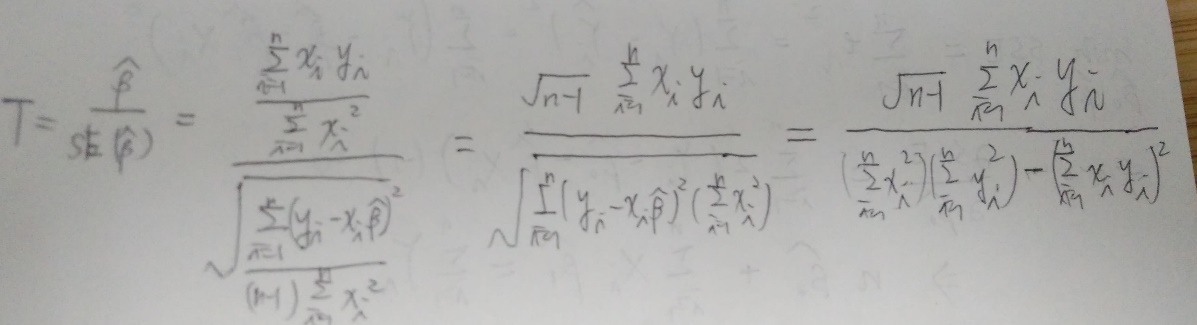
(B)

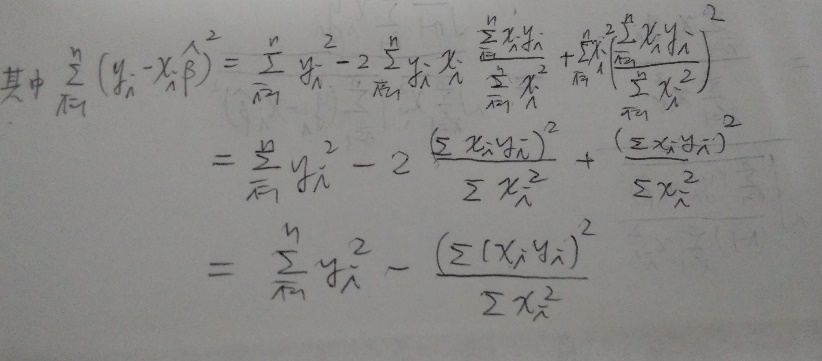


根據p-value值非常的小,我們有顯著證據拒絕H0。

(C)

我們從上圖可以看出把自變量和應變量互換去做線性回歸,得出來的P-VALUE和t test 的值是一樣的。

(D)

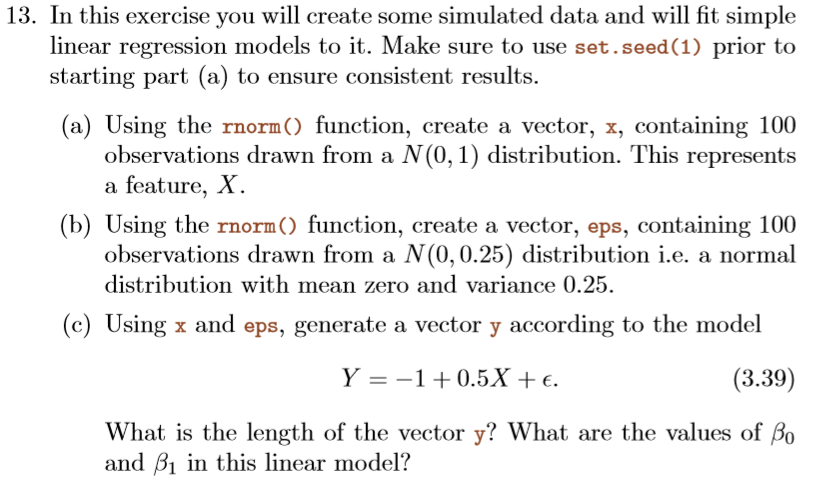


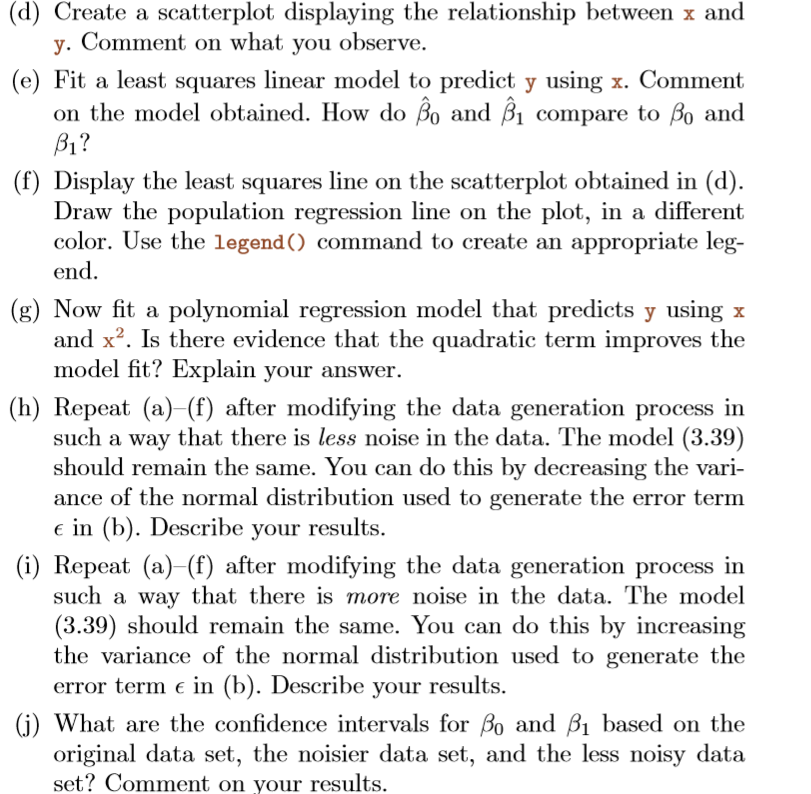
(e)

由(d)答案可知 f(x,y)=f(y,x)

(f)

由(a) , (b)可知有相同的t-test的檢定結果。





(a)



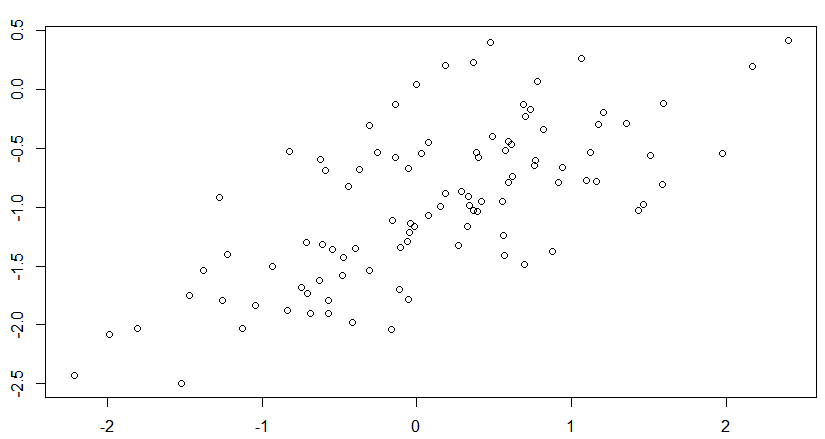
(b)



(c)

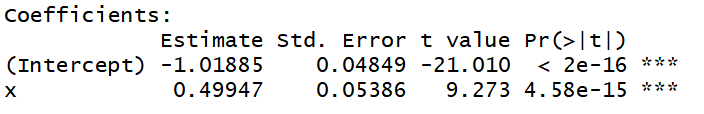
y長度100,β0=-1 , β1=0.5

(d)



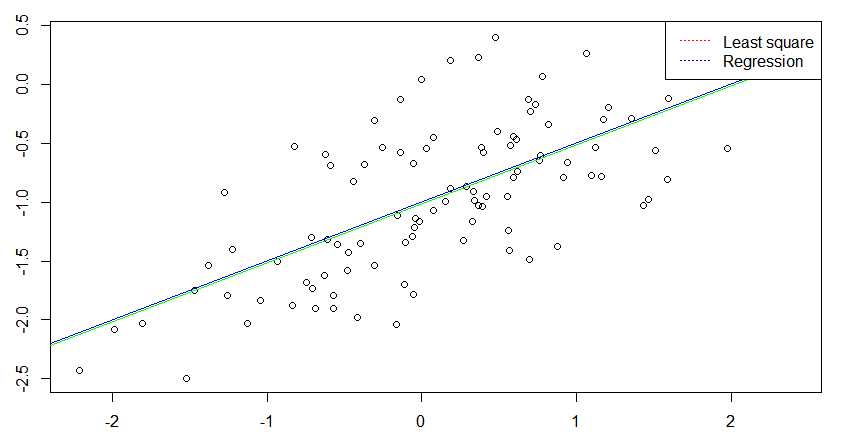
X和y有線性的趨勢。

(e)

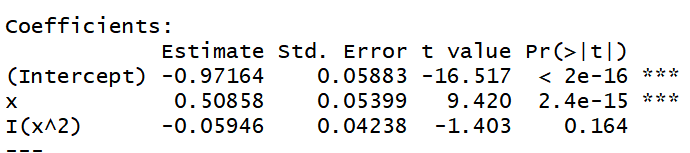


=-1.01885 , =049947 ,這跟β0=-1 , β1=0.5非常接近,且p-value非常小,我們有顯著證據拒絕H0。

(f)

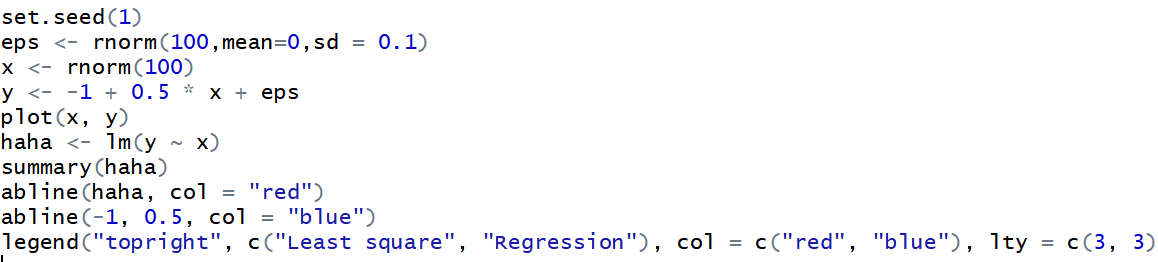


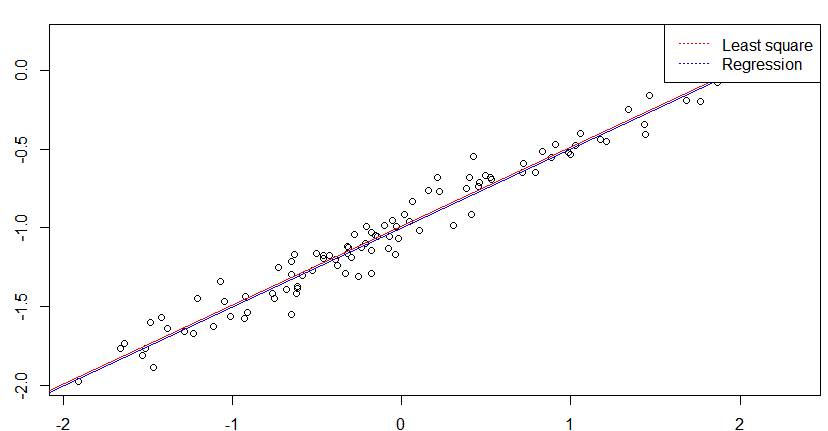
(g)



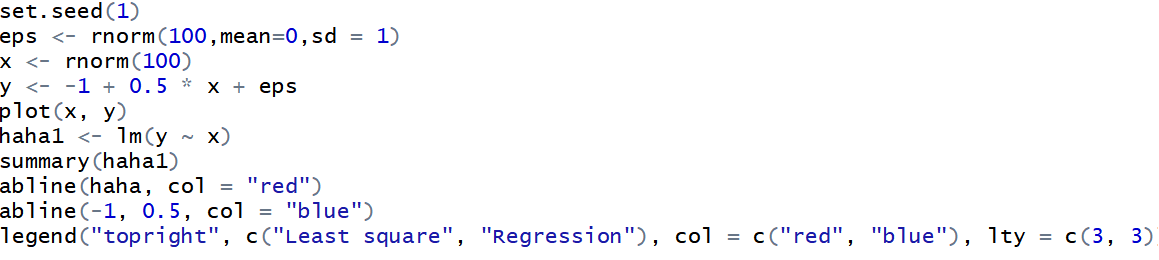
2次項的p-value=0.164,所以我們沒有顯著證據reject H0。有二次項的回歸方程式雖然R^2會比簡單線性回歸高,但其實回歸的解釋資料能力沒有比較好。

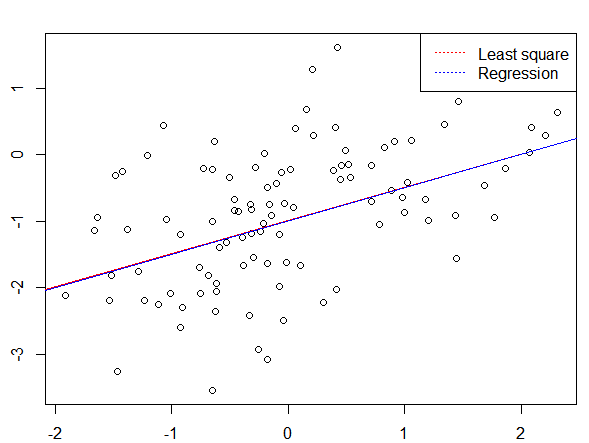
(h)



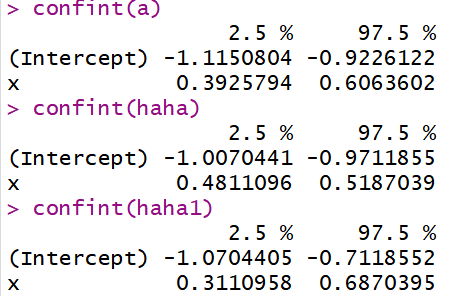


(i)





(j)



隨著noise越多,信賴區間越寬,反之;noise越少,信賴區間越窄,代表有比較好的預測。