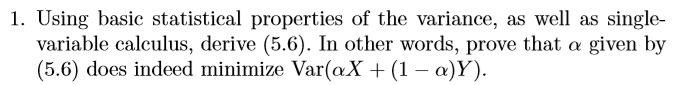
M072040019 梅瀚中

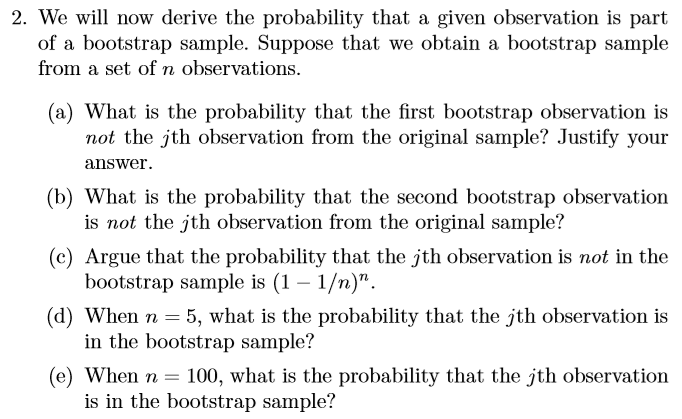


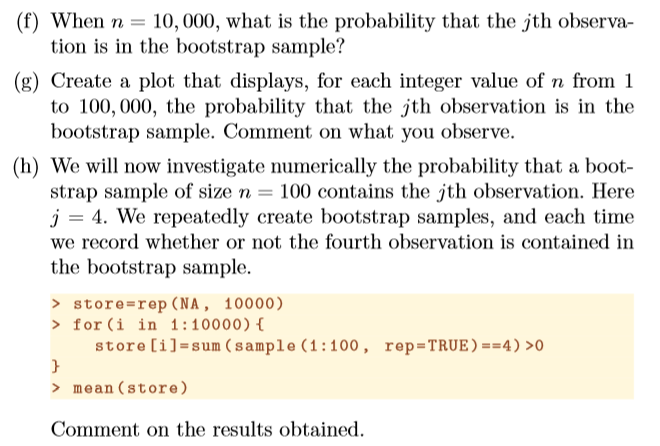


 =0









(a)

1-1/n

(b)

1-1/n

(c)

題意=在n次裡的boostrap sample,每一個boostrap observation都不是第j個observation的值

(1-1/n)\* (1-1/n)\* (1-1/n)\* (1-1/n)\*…….\* (1-1/n)= (1-1/n)^n

(d)

P(全部- jth observation is not in the bootstrap sample)

=1- (1-1/5)^5

=0.67232

(e)

P(全部- jth observation is not in the bootstrap sample)

=1-(1/100)^100

= 0.6339677

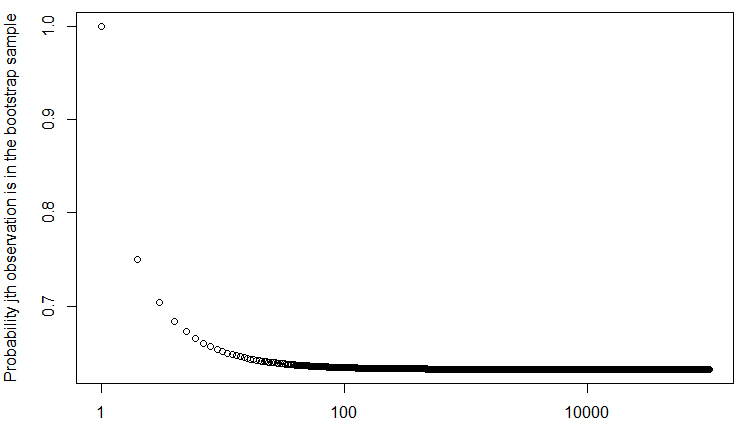
(f)

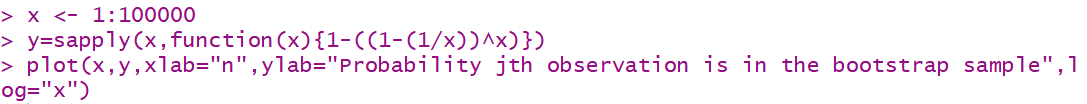
P(全部- jth observation is not in the bootstrap sample)

=1-(1/10000)^10000

=0.632139

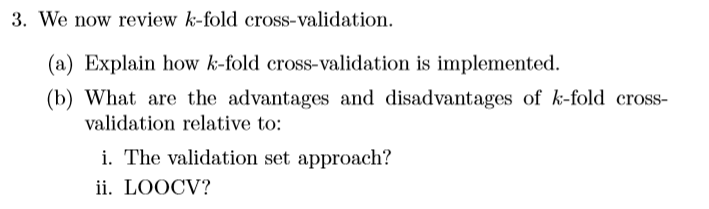
(g)





(h)

我們得到從boostrap sample size n,抽到第j個觀察樣本的機率,最後會收斂到0.632。



(a)

原始的樣本數據被隨機分成k份,在k份子集合中每次挑選一份作為vaildation set,其他k-1份作為training set, 在每一個training set 上訓練後得到一個模型,用這個模型在testing set上測試。這樣的過程重複進行k次,每一份都會有一次作為vaildation set。最後在把這k次的測試結果進行平均,得到最後的測試結果。

(b)

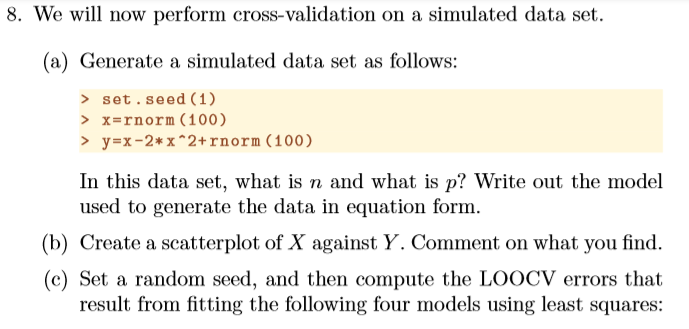
(i) 缺點:相較k-fold cross validation ,vaildation set 法的test error會有較高的變異程度

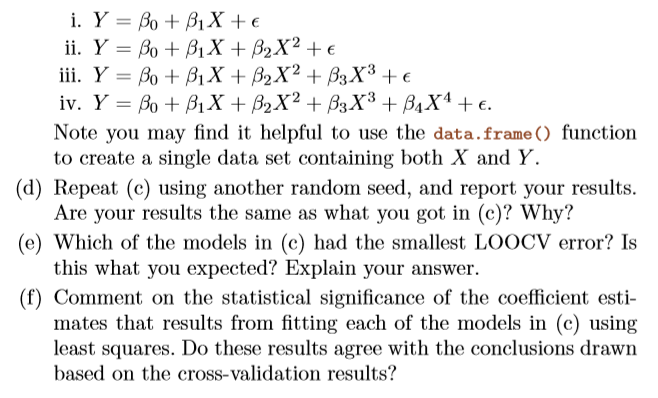
優點: 相較k-fold cross validation ,vaildation set法,節省運算時間很多。

(ii) LOOCV法就是k-fold cross validation,只是他的k=樣本數

缺點: 如果樣本數很大,相較k-fold cross validation,LOOCV法運算時間增加非常多。

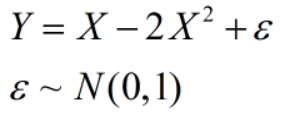
優點:LOOCV, 每一回合中幾乎使用所有的樣本訓練，因此最接近母体樣本的分布。



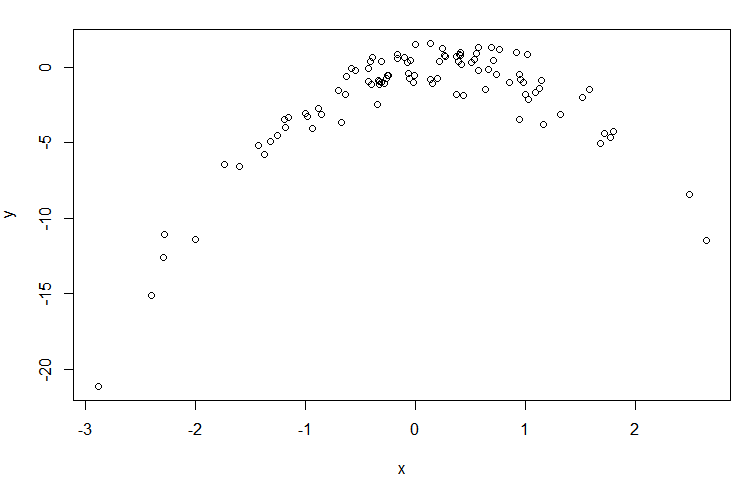


(a)

n=100 and p=2

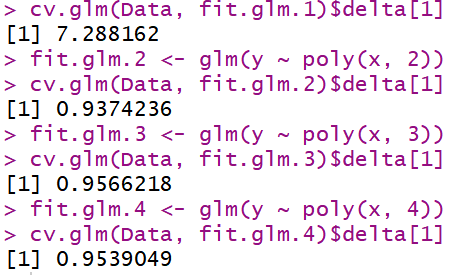


(b)

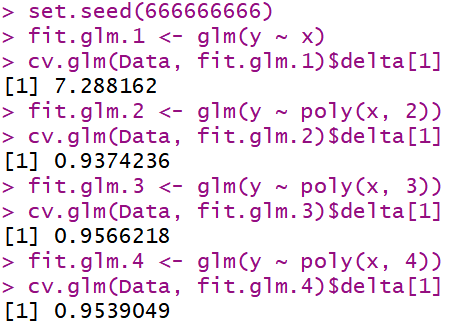


呈現非線性相關。

(c)



(d)

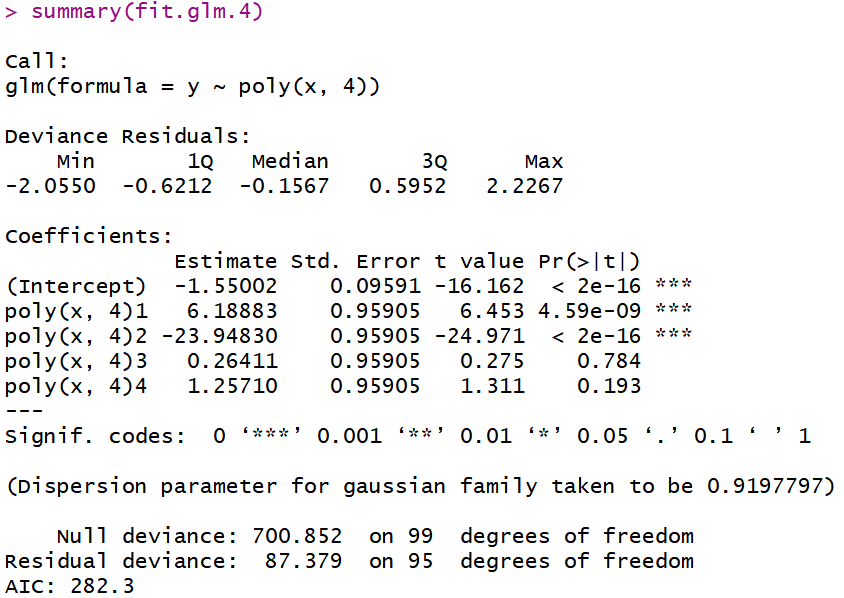


因為LOOCV是把每一筆的資料都會當作test,所以不管seed是從哪一筆觀察值開始,結果都會一樣。

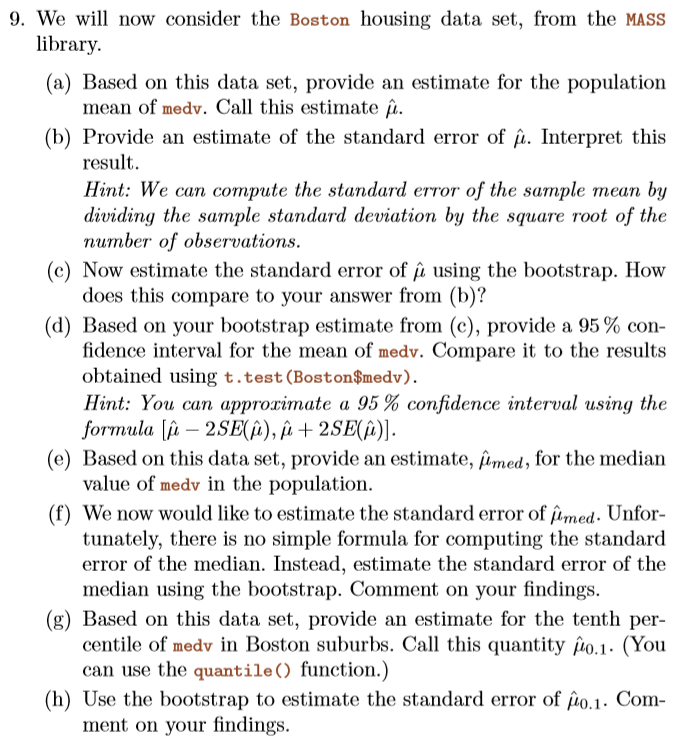
(e)

Fit.glm2的test MSE 是最小的。

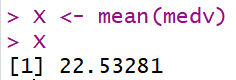
(F)



只有線性和二次項的係數是顯著的,跟我們用交叉驗證得出的結果是一致的

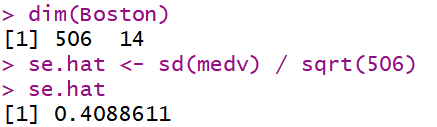


(a)

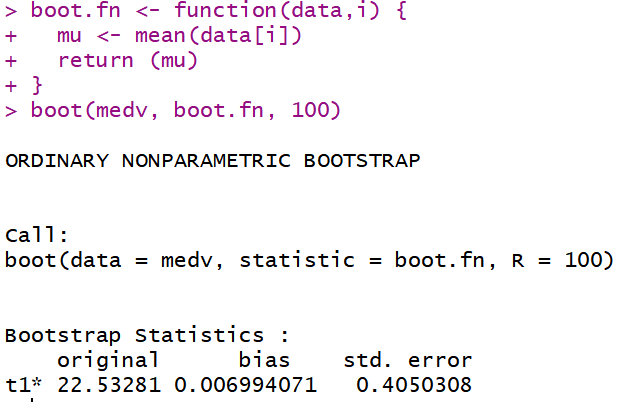


(b)



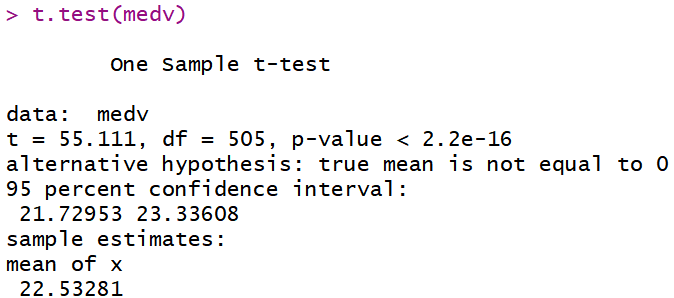


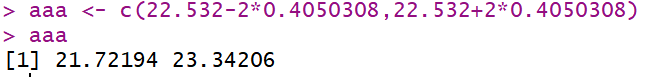
(c)



Boostrap:0.40530308

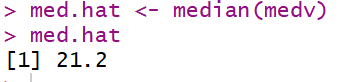
(d)



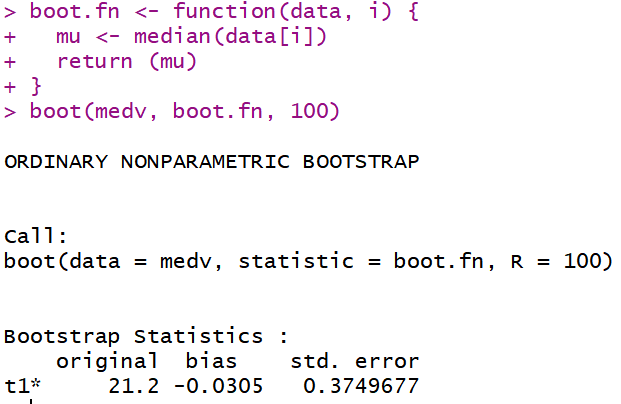


兩個信賴區間 差不多。

(e)

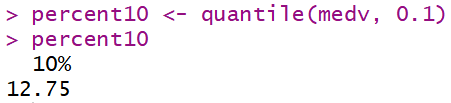


(f)



兩個得到的中位數估計值很接近。

(g)



(h)

