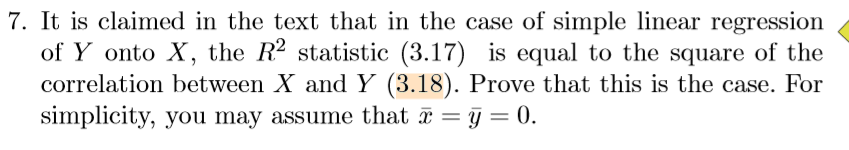
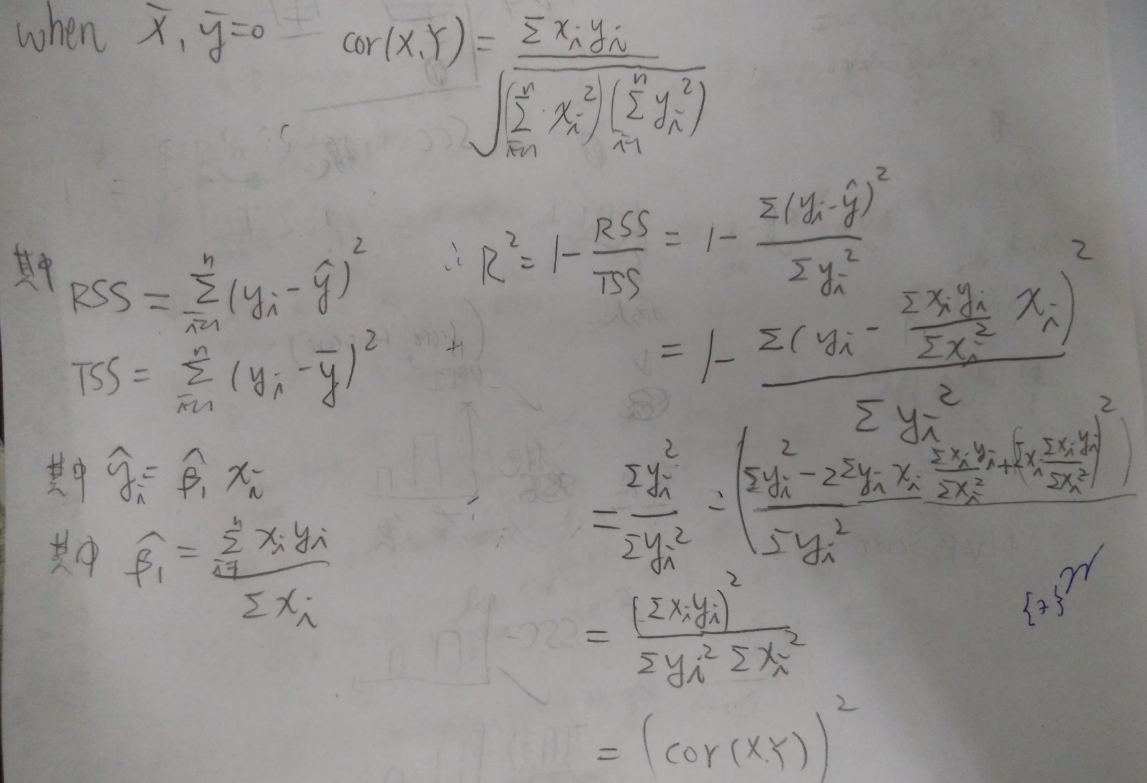
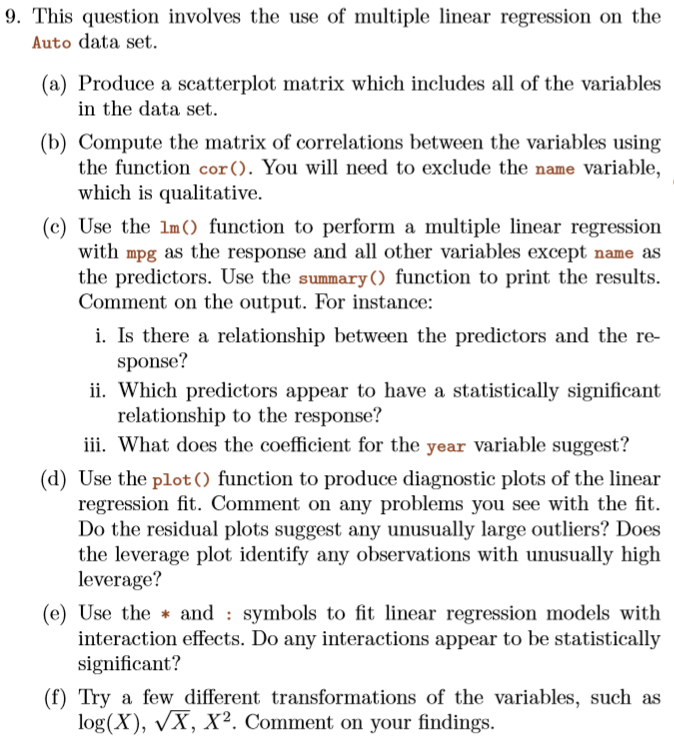


Knn classifier是一個分類的演算法,演算法的核心思想是如果一個樣本在空間中的K個最相鄰的樣本中的大多數屬於某一個類別,則該樣本也屬於這個類別

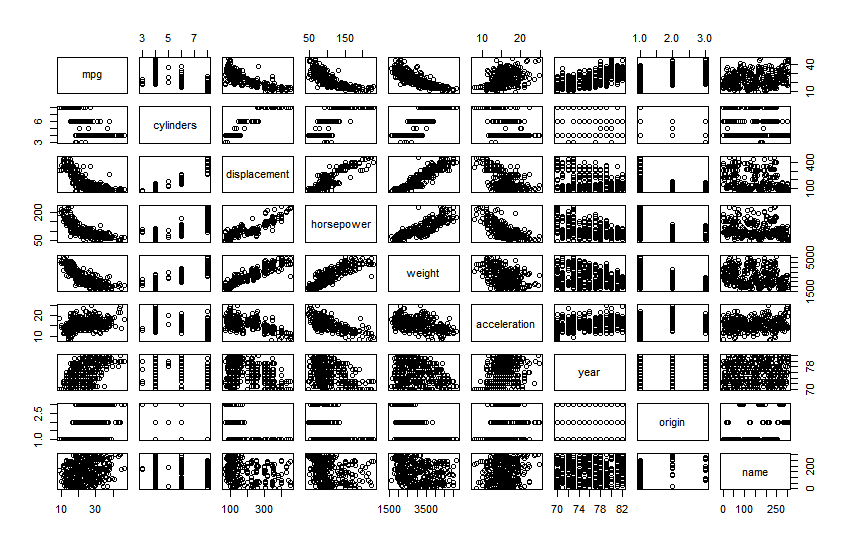
Knn regression 跟一般的regression一樣,應變數是屬於連續變數,想法是,一個樣本在空間中的k個最相臨的樣本的值(y),作平均,得到的值為此樣本的值。



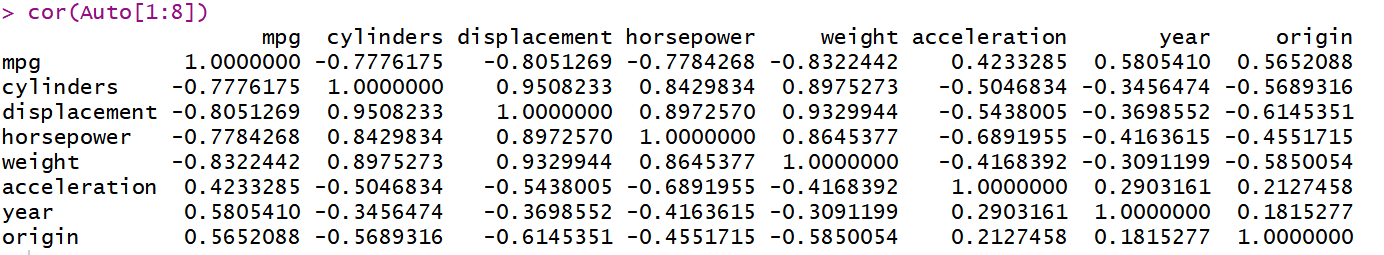




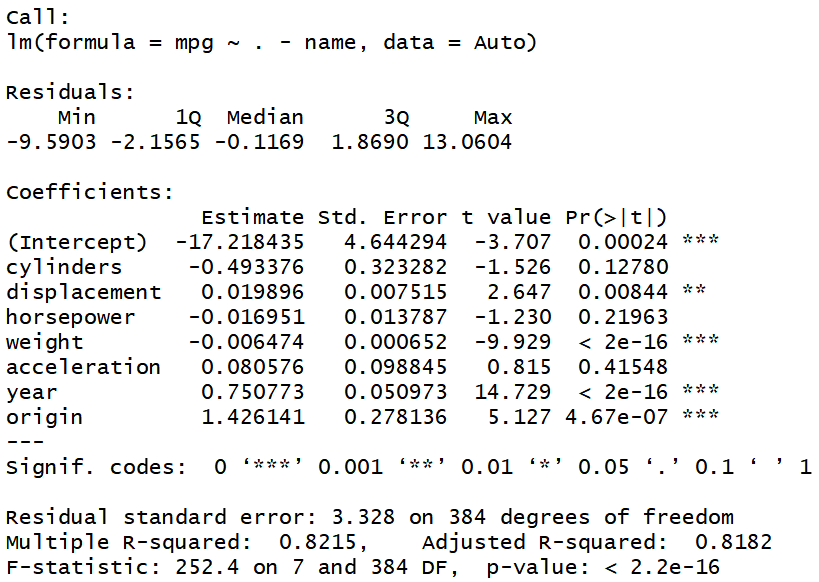
(a)



(b)



(c)

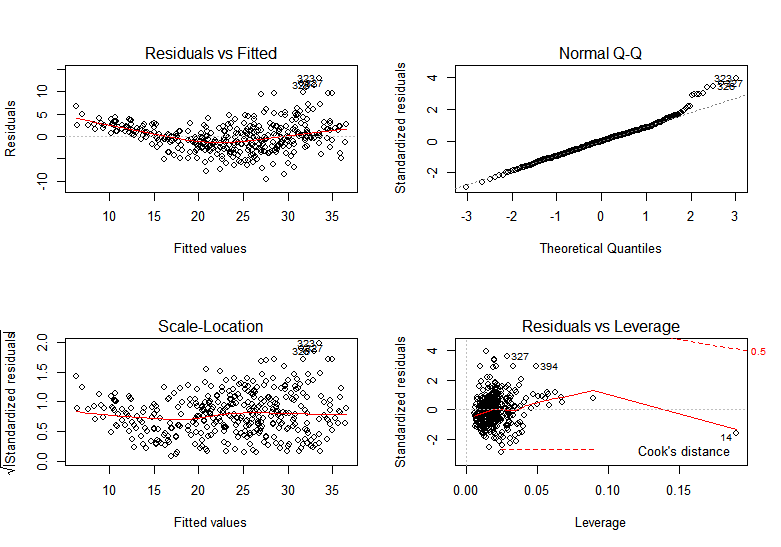


(i)p-value<2.2e-16,所以整體來說y跟x有關係。

(ii)displacement,weight,year和origin變數都和mpg有顯著關係。

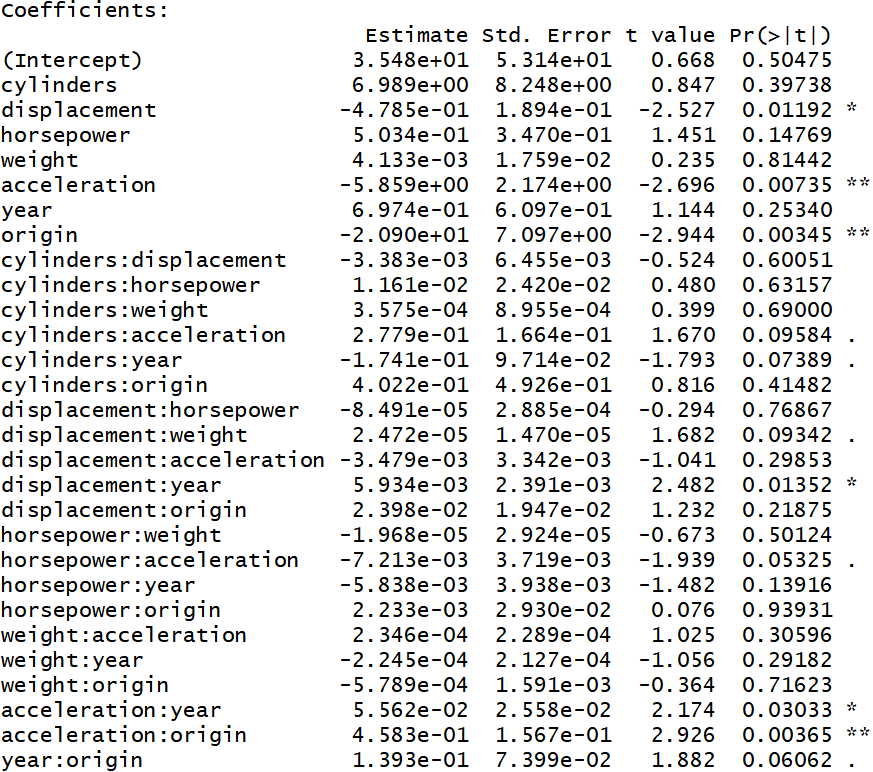
(III)平均來說,year增加一單位,mpg會增加0.750773單位

(d)



我們看左下方的圖,圖中左上方的幾個點的standardized residuals有一點大,但是還沒有明確的超過範圍。右下方的圖中我們可以看到點14是一個高的槓桿值。

(e)

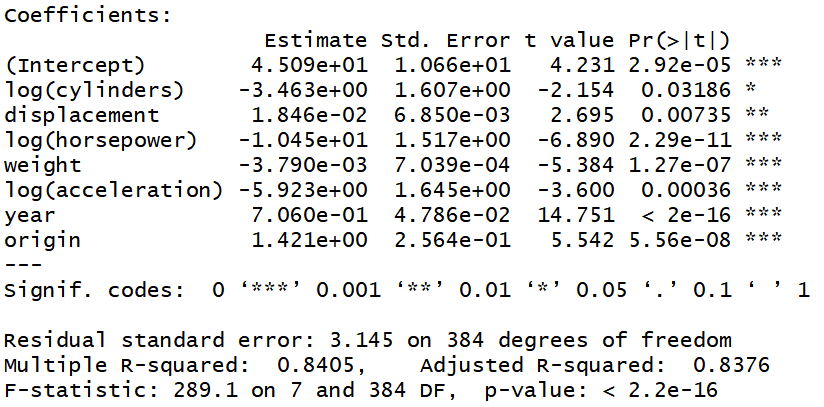


Displacement和year的交互作用項有在alpha設定為0.05之下有顯著效應,

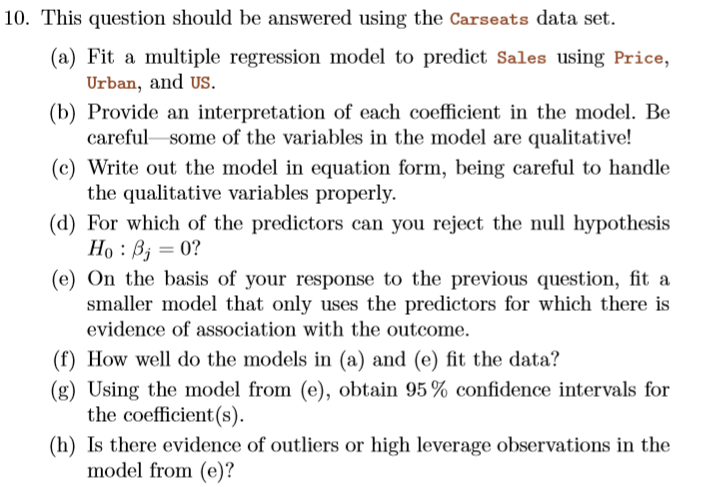
Acceleration和year的交互作用項有在alpha設定為0.05之下有顯著效應,

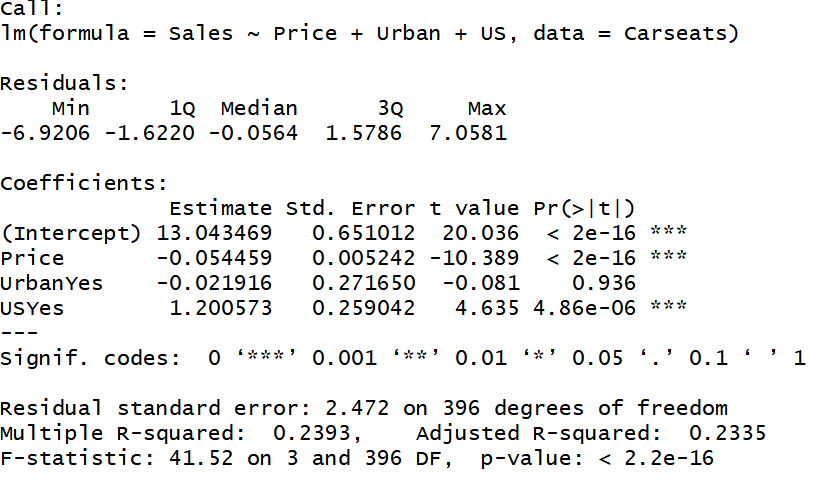
Acceleration和origin的交互作用項有顯著效應。

(f)



把(c)原本不顯著的變數加上log後,變數變成顯著。



(a)

(b)

Price: 公司對每個站點的汽車座椅收費

Urban: 表示商店是在城市還是鄉村

US: 表示商店是否在美國

(C)

Sales=13.043469-0.054459 price-0.021916 Urban+1.200573 US

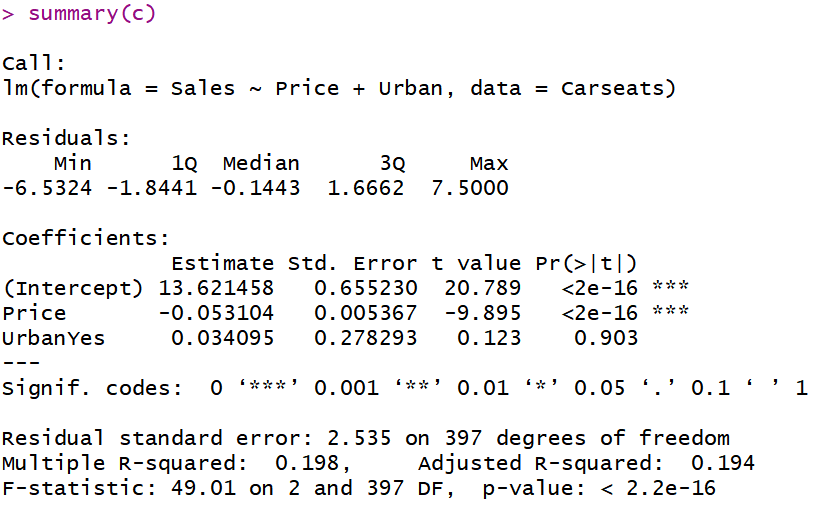
(D)

Price的p-value<2e-16 ,

US的P-value的4.86e-06

這兩個變數reject H0

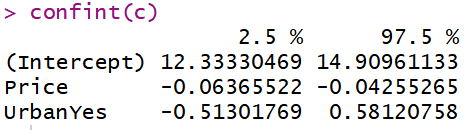
(e)



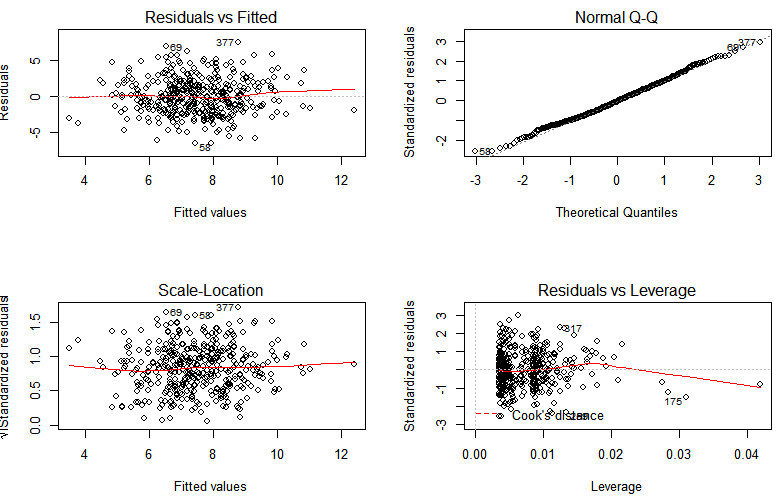
(F)

只用price,Urban比用price,Urban,US的R^2小,代表此模型解釋能力不夠。

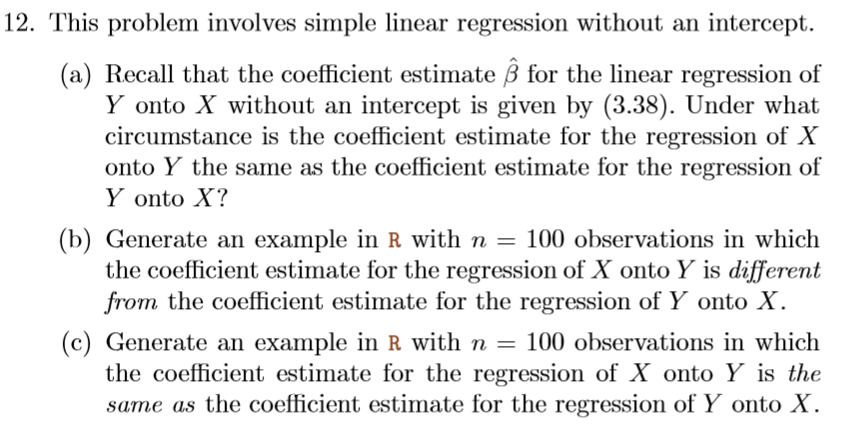
(G)

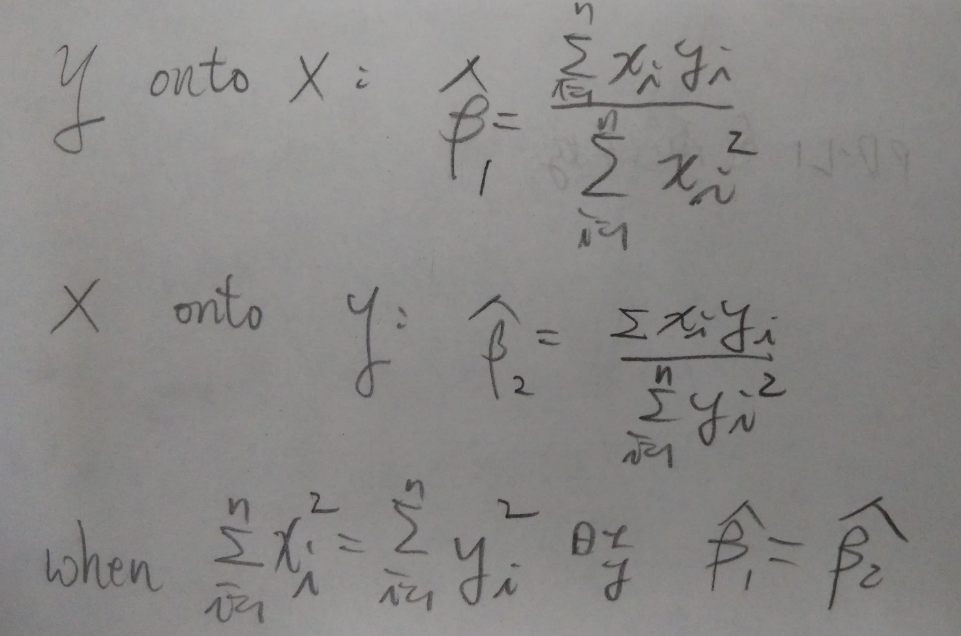


(h)

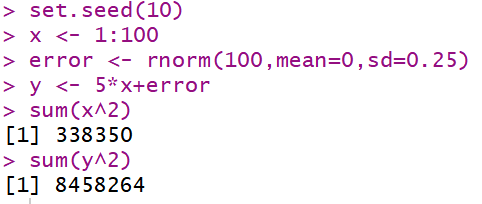


點377,69 ,58 都是outliers。點175是high leverage



(a)

(b) 查看無截距模型的beta係數時,X和Y互換,只需要看分母值相不相同:



這裡sum(x^2)和sum(y^2)不同 所以估計的beta就不相同

(c)

把error設為0

