Statistical Learning and Data mining

Homework 3

M052040003 鍾冠毅

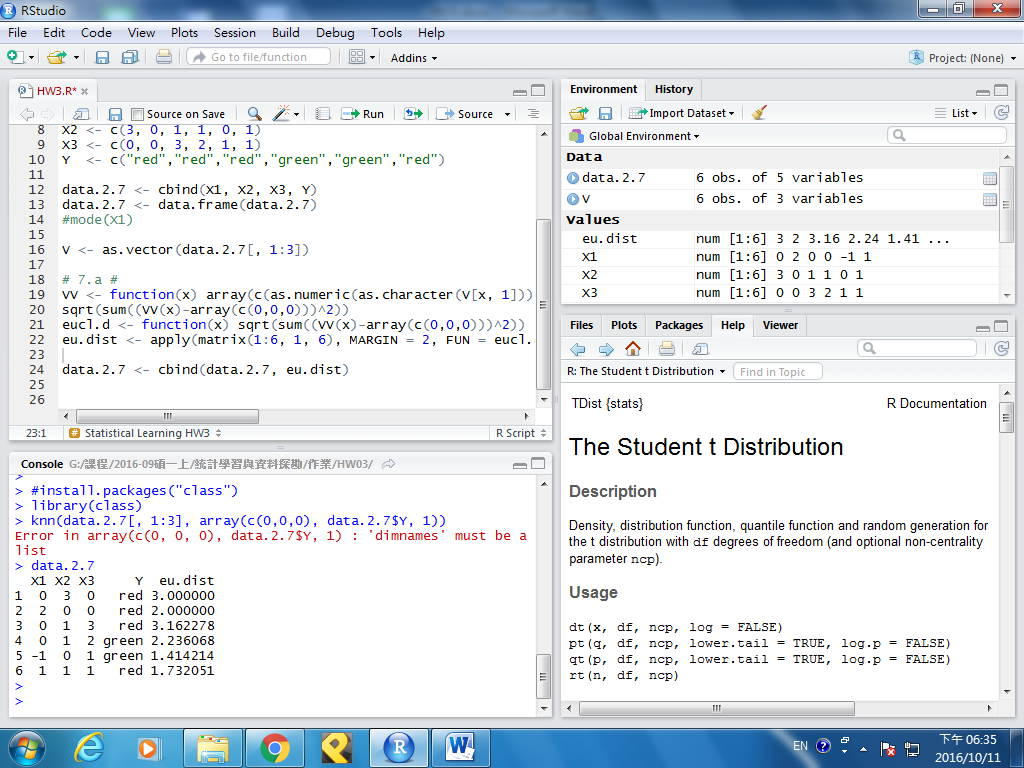
Chapter 2

3.a.

3.b. Bias與training error 隨著彈性增加而減少，為嚴格遞減；variance隨著彈性增加 而增加，為嚴格遞增；test error則是凹口向上，也就是說有極小值，而我們在尋找模型的方向也是期望得到越小的test error越好；Bayes error用來檢測classification model，其值介於0、1之間，係因為其公式為

Bayes error也被稱為irreducible erroe，可以說是模型誤差的最大下界。

7.a.



7.b. green，最靠近（0, 0, 0）的一個點為綠色。

7.c. red，最靠近（0, 0, 0）的三個點有二紅一綠，則分類預測為紅色。

7.d. small，非線性可用較少的K以得到較高的彈性。

Chapter 3

1. 三個虛無假設皆設定該變數之係數為0，也就是說在其他兩個變數固定的狀況下，該變數對模型沒有影響。從p-value的大小可以知道其顯著性，三個變數中TV與radio之值<0.0001，拒絕虛無假設，意即兩變數之係數不為0；反之newspaper的p-value大於0.05，表示沒有足夠證據拒絕虛無假設，則其係數為0。

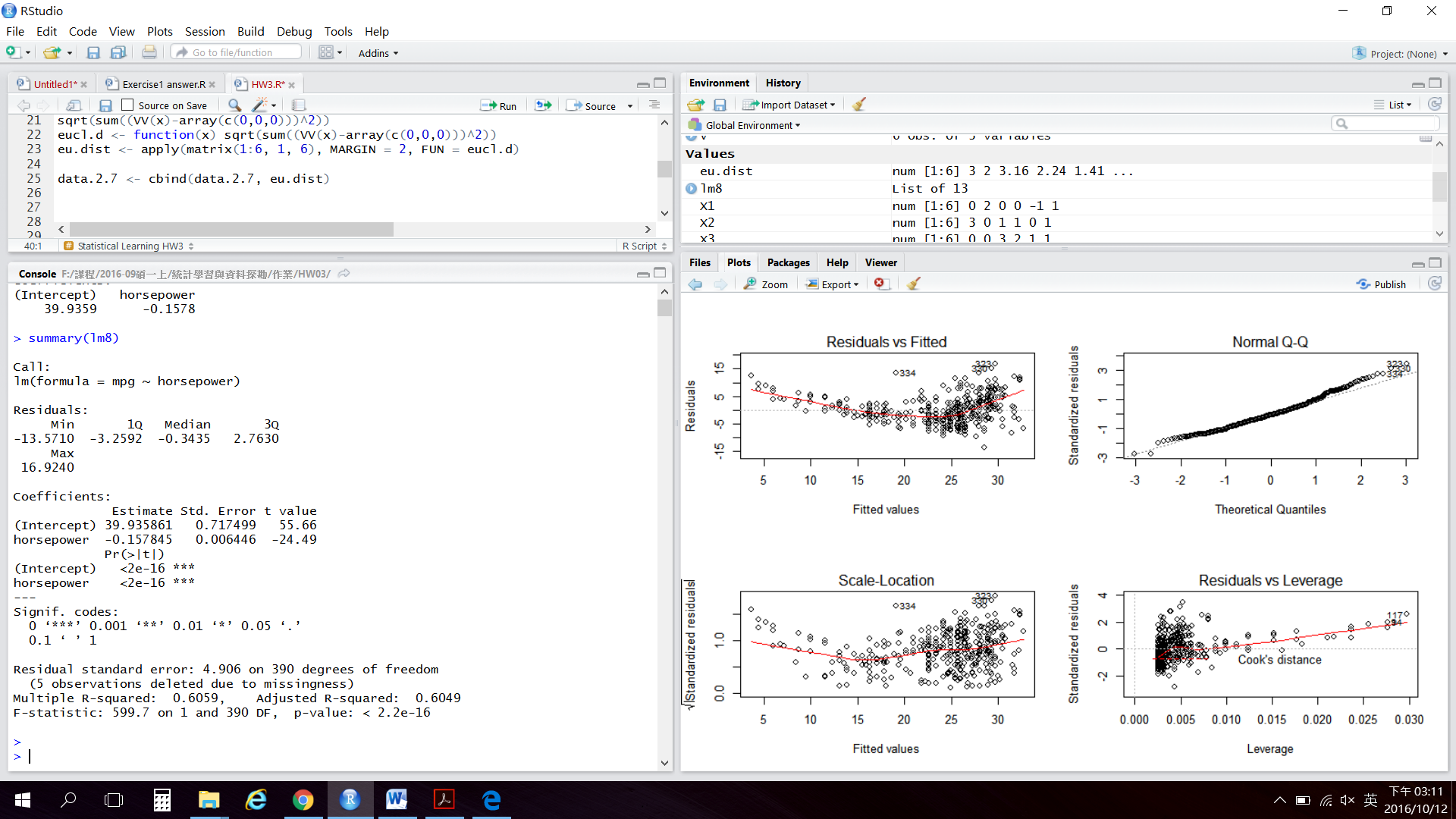
3.a.

因為GPA與Gender交叉項的關係，以Gender為變數之係數經整理可以得到，也就是說此項需知道GPA固定值為多少，方能知道性別對薪水的影響。因此GPA夠高（>3.5）的話Gender的係數，則男性將獲得更高的報酬。iii

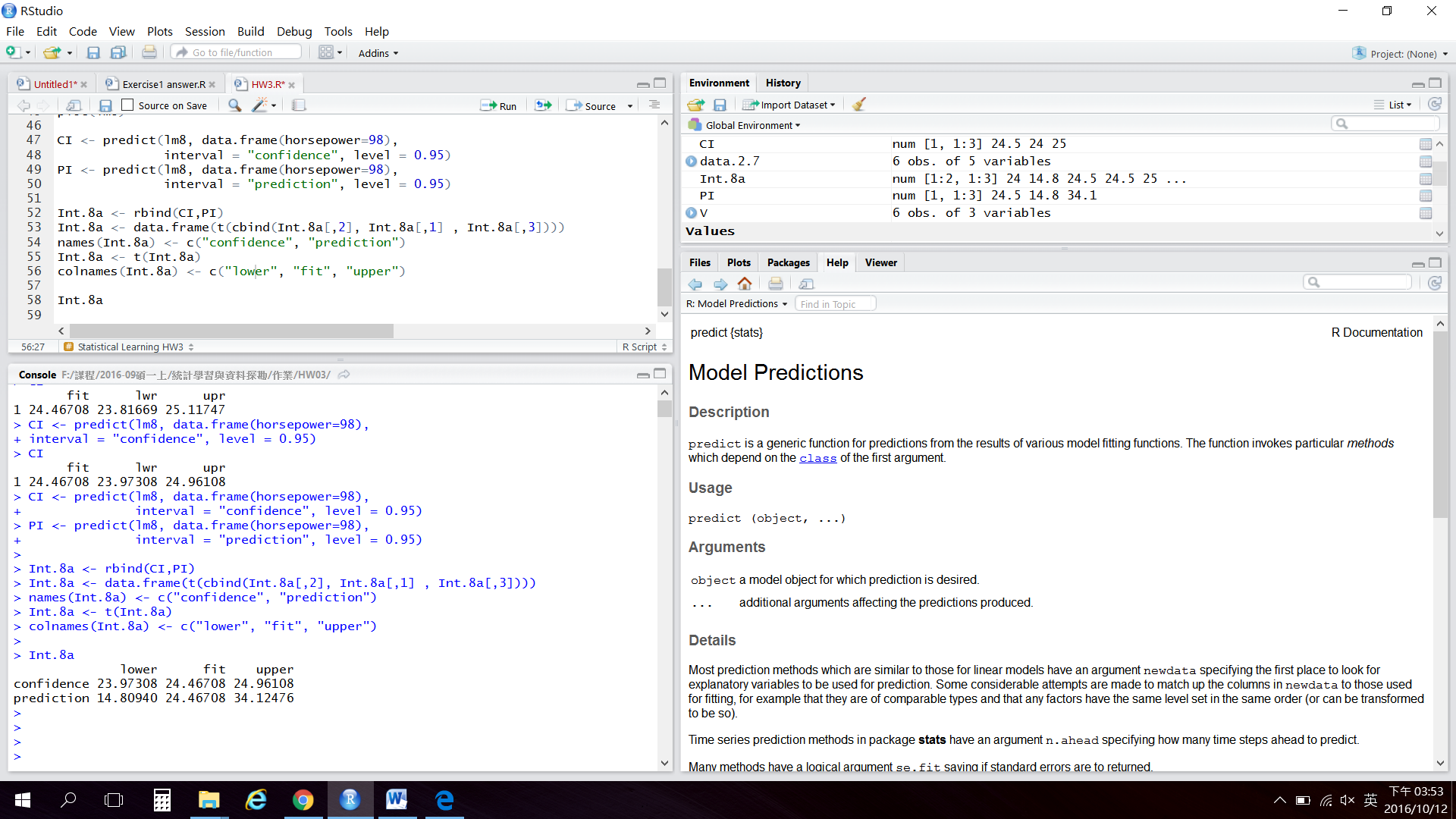
3.b.

3.c. False。雖然係數的估計值很，但是其standard error亦可能很小，兩者相除可能有不一樣的結果，故需檢查各係數之p-value，來決定該變數是否影響反應變數。

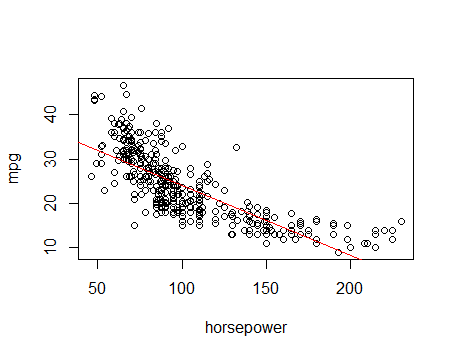
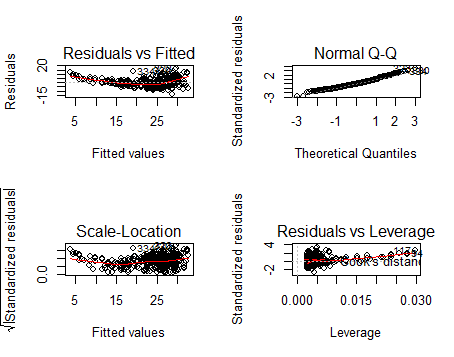
8.a.



1. , so the relationship is strong.
2. For the coefficient of horsepower is negative, the relationship is negative as well.



8.b. 8.c.

兩筆資料可能成非線性。