**111學年度第二學期科學計算軟體作業八**

系級:測量系114 姓名:黃薇庭 學號:F64101032

1. 讀取HW\_Database.csv資料集，內含2020/04/01 – 2020/05/09共39日之臺南空氣品質監測站PM2.5、NO2、Temperature、RH數值，試以線性迴歸說明PM2.5、Temperature及RH對NO2之影響效應，內容需含下列項目(**20%，**答題提醒: **注意標註p值**並說明是否達統計之顯著水準，若未達到或錯誤皆會斟酌扣分)：  
   \*小提醒：資料集中PM2.5變數名稱為PM25
2. 各項因子之**Pearson相關檢定**(說明各項因子與NO2之關係以及其是否具統計上之意義)。  
   \*選定資料集中特定欄位的方式: Dataset[5:8] (選擇Dataset中5-8欄進行作業)。  
   \*也可以刪除不需要的欄位之後再做分析。
3. 以**ANOVA**方式探討該線性模型之**整體配適度**如何，**是否達統計上之顯著水準，****模式中有無影響NO2之因子(注意若使用anova\_alt需要先執行定義函數的語法，於下兩頁，**source: https://community.rstudio.com/t/anova-table-for-full-linear-model/42074/10 created by RussS**)。**
4. 根據模式輸出之結果報表，說明**哪幾項因子(不含常數項/截距)對於NO2之影響具統計上之顯著意義**，整體模式之**解釋能力(決定係數)為何**？
5. 列出**完整模型公式(均四捨五入至小數點第二位)**。
6. 模型**校正後決定係數值**，及其計算**公式**為何(需詳列公式，原始R2四捨五入至小數點第二位後才帶入，結果也四捨五入至小數點第二位)？

ANS:

由下圖可以發現，

NO2與PM25為高度正相關(r=0.75)，然而p-value=0.0000<0.05，達到統計上之顯著水準

NO2與Temperature為負相關（r=-0.57），然而p-value=0.0002<0.05，達到統計上之顯著水準

NO2與RH為負相關(r=-0.24)，然而p-value=0.1455>0.05，未達到統計上之顯著水準

一張含有 文字, 螢幕擷取畫面, 字型, 收據 的圖片

自動產生的描述

v

v

v

v

一張含有 文字, 字型, 螢幕擷取畫面, 收據 的圖片

自動產生的描述

p-value = 4.5234e-7 < 0.05達顯著水準，表示整體模型至少有一項因子對於 NO2 有影響。

一張含有 文字, 螢幕擷取畫面, 字型, 代數 的圖片

自動產生的描述

Temperature 則為 p = 0.000174 < 0.05 達顯著水準，表示 Temperature 對 NO2 有影響。

一張含有 文字, 字型, 螢幕擷取畫面, 白色 的圖片

自動產生的描述

PM25則為 p = 3.455e-8 < 0.05 達顯著水準，表示PM25對NO2有影響。

一張含有 文字, 字型, 螢幕擷取畫面, 白色 的圖片

自動產生的描述

RH 則為 p = 0.1455 > 0.05 未達顯著水準,表示 RH 對NO2沒有影響。

一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述

根據T檢定結果之 p-value，可知截距 p = 0.235 >0.05 未達顯著水準，截距為 7.27225。

PM25則為 p = 5.03e-5 < 0.05 達顯著水準,對於 NO2 的影響具有統計上之意義，PM25之迴歸係數＝0.28689

Temperature 則為 p = 0.197 >0.05 未達顯著水準，對於 NO2之影響沒有統計上之意義，其迴歸係數 = -0.17109

RH 則為 p = 0.461 > 0.05 未達顯著水準，對於NO2之影響沒有統計上之意義，RH 之迴歸係數 =0.03474

整體模式對於 NO2 之決定係數 = 0.5979，能夠解釋NO2變化的 59.79%，p-value = 4.523e-7 < 0.05，達統計上之顯著水準,是為統計上可信之結果。

模型公式:

NO2(y) = 7.27225 + 0.28689 ×PM25 -0.17109 × Temperature +0.03474 ×RH

校正後決定係數:

一張含有 文字, 收據, 字型, 文件 的圖片

自動產生的描述

#code:

一張含有 文字, 螢幕擷取畫面, 字型, 文件 的圖片

自動產生的描述