統計諮詢 HW1

RE6091020 汪玄同

Problem 4.2

資料為Sokal and Rohlf (1981) 與Hand et al. (1994)從41個城鎮收集的空汙檢測資料，每筆資料共有7個變數，分別為:

SO2: 空氣中的SO2含量，每立方米mcg

temp: 年平均溫度，以華氏度為單位

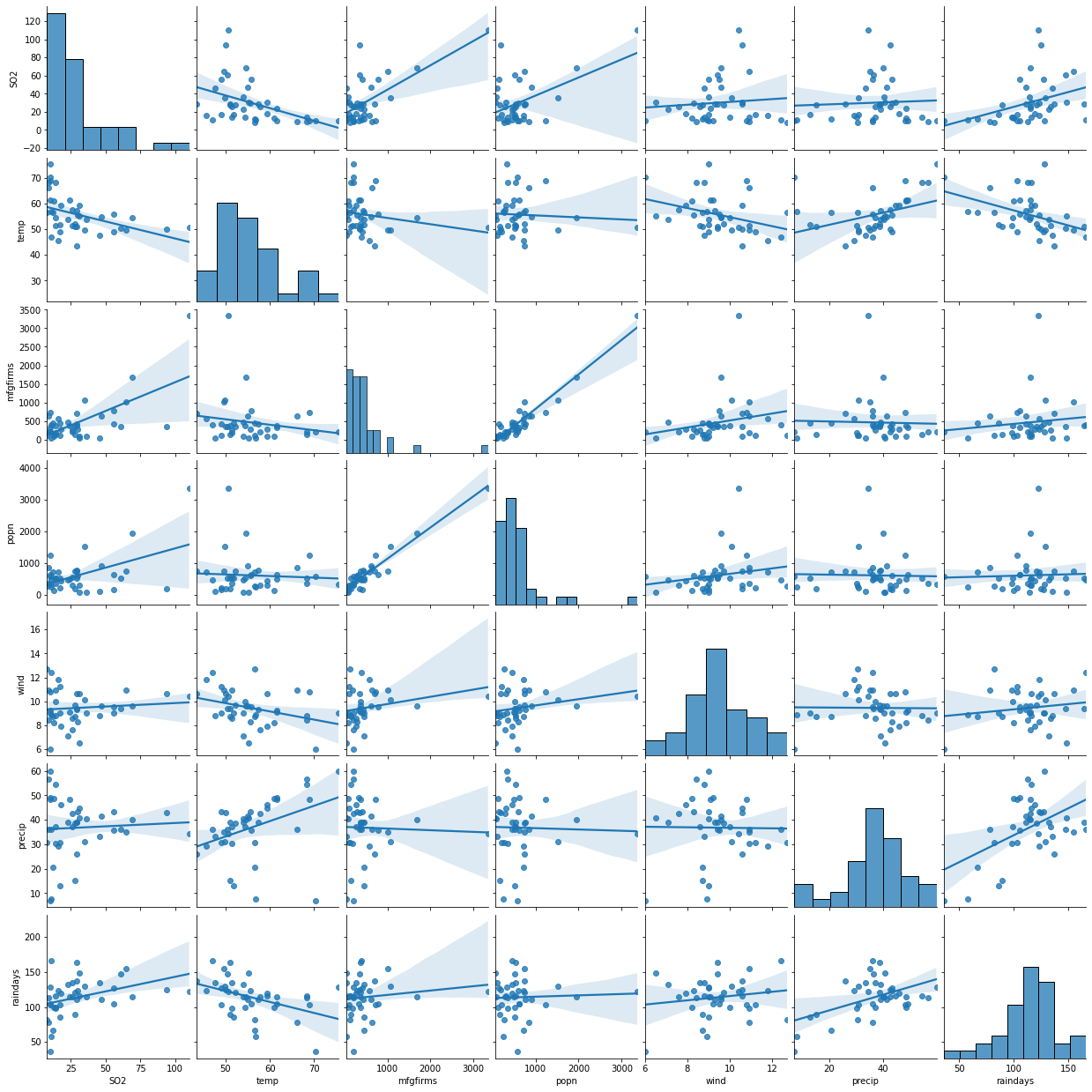
mfgfirms: 僱用至少20名工人的製造公司的數量

popn: 1970年人口普查人口規模

wind: 年平均風速（英里/小時）

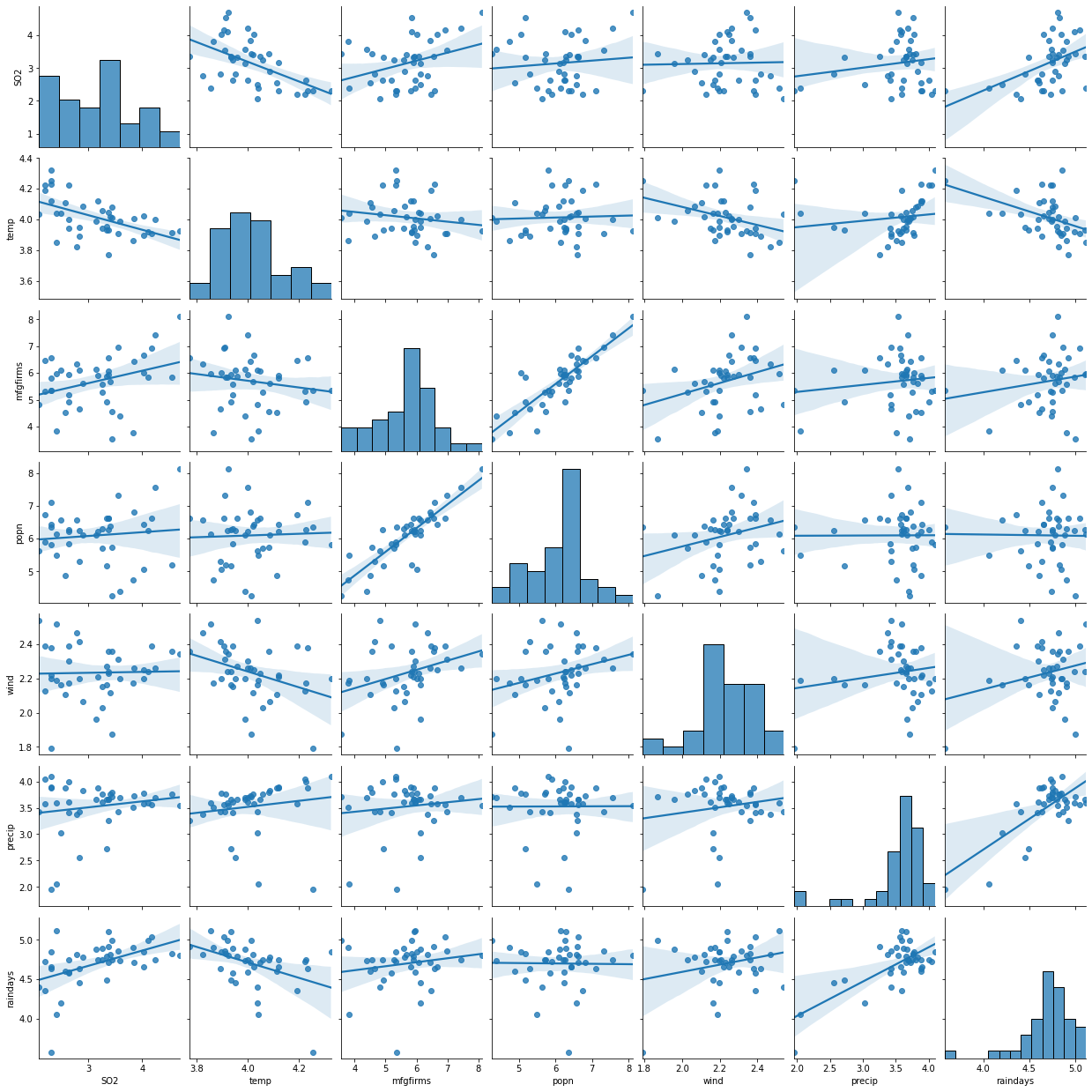
precip: 年平均降水量（英寸）

raindays: 每年有降水的平均天數

Scatter plot matrix:

由於原本資料中的數據有許多離群值，使某些圖的點集中在某一小區塊，並且回歸線受那些點影響太大，使用log進行轉換可以使點的差距減小，點較集中回歸線較不會受到離群值影響太大。

Log scatter matrix:



變數SO2與其他六個變數的相關係數(經過log轉換後的資料):

Correlation SO2, temp: -0.53

Correlation SO2, mfgfirms: 0.33

Correlation SO2, popn: 0.10

Correlation SO2, wind: 0.02

Correlation SO2, precip: 0.17

Correlation SO2, raindays: 0.47

經過絕對值轉換後與SO2擁有最高相關性的變數為temp，相關係數為 -0.53

Problem 5.13

(a)

資料為脊椎受傷的患者治療前與治療後的脊椎角度量，由資料描述可知每筆資料為同一病人治療前後的脊椎角度值，兩個數值並非獨立，使用成對樣本t檢定。

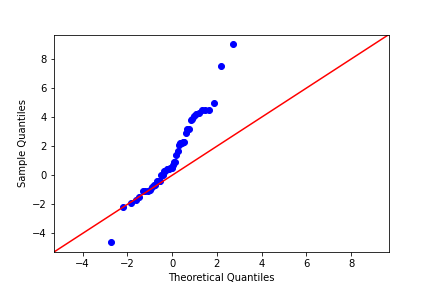
先使用適合度檢定測試資料是否為常態，假設新變數D為治療前檢治療後的隨機變數D = Pre – Post，Pre為治療前變數，Post為治療後變數。

檢定統計量 0.9600,，p-value = 0.1012，在95%的信心水準之下，我們沒有充分的證據顯示資料不是常態分佈。

假設相減的資料為常態分佈，使用成對樣本t檢定來檢定D是否小於等於0

統計量為t = 3.556，p-value=0.0004，在95%的信心水準之下，我們有充分的證據顯示D變數大於於零，代表治療後的角度是顯著小於治療前的角度。

(b)

自由度為5的t分配與治療後角度的QQ Plot:

在qq-plot中若資料符合分配的話資料形成的線應該要擬合斜率為1並經過原點的線，但這張圖資料與線並未擬合，判斷資料並不服從自由度5的t分配。