

1. Write down your **regression equation** in basic part (1%)

For city A:

$$y_i = 30.672 - 0.108(0.6x^2 - 18x - 45) + 0.008(0.6(y_{i-1})^2 - 18y_{i-1} - 45)$$

For city B:

$$y_i = 27.178 - 0.04(0.8x^2 - 14.6x - 114) + 0.011(0.8(y_{i-1})^2 - 14.6y_{i-1} - 114)$$

For city C:

$$y_i = 37.249 - 0.00033(0.8x^2 - 134x + 7486) - 0.004(0.8(y_{i-1})^2 - 134y_{i-1} + 7486)$$

2. Briefly describe the **variables** you used in the advanced part (1%)

除了在 basic part 就有用的 variable(溫度和前一天的登革熱確診人數)之外，我在 advanced part 還多考慮了各個城市的降雨量。

3. Briefly describe the difficulty you encountered (1%)

在這次的 assignment 中，我主要在兩個環節遇到困難，分別是在資料前處理和建立整個 regression equation 這兩部分。前者是因為 python 不夠熟練，以及 outlier 的排除，後者則是為了要到一個比較好的模型。

4. Summarize how you solve the difficulty and your reflections (2%)

在 python 不熟練這個問題，解決的方法就是多查資料，想要用什麼功能都 google。而在處理 outlier 上，我是計算整個 column 的資料，求初期標準差和平均值，然後求每天溫度的標準化分數(Z score)，將分數大於或小於 1 的資料，整個 row 剷除，用此方式來清掉 outlier。在建 regression equation 上，我到最後也沒有找到一個比較好的流程來尋找 equation，只能花時間，透過想像力，並且觀察每次修改後的情形。對於這次作業我覺得難度還算是適中，雖然拿高分可能有點難(因為要和其他同學競爭)，但是要拿到基本分應該不能(吧?)。所以我覺得這份作業寫起來挺有趣的，有讓自己更了解課堂中的內容。