

HW2:PM2.5 分析及預測

黃敏嘉、郭宇鑫

GOAL: 根據 PM2.5 的相關因子預測 PM2.5 的值

step 1: 選取測站，選取跟 PM2.5 可能的相關變數

step 2: 根據 p 值篩選變數

step 3: 根據所分析資料的模型，預測 PM2.5 值

step 1: 選取測站，選取跟 PM2.5 可能的相關變數

我們選取的 PM2.5 測站是台中的忠明測站，並且下載 2015 年第 1~330 天，每一天的 PM2.5 值，相關變數則選擇前一天的 CO、NO、NO2、O3、SO2、相對溼度、風速，再加上前一天同一測站的 PM2.5 作分析 (最後用分析結果預測第 331 天以後的 PM2.5)，迴歸模型為

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_8 x_8 + \epsilon$$

step 2: 篩選變數

先把所有變數丟進去做分析，結果如下：

```
##  
## Call:  
## lm(formula = y ~ x.pre + x.CO + x.NO + x.NO2 + x.O3 + x.SO2 +  
##      x.RH + x.WS)  
##  
## Residuals:  
##      Min        1Q    Median        3Q       Max  
## -33.232  -5.457  -0.957   4.357  43.255  
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.31603    9.27764   0.896 0.370747
## x.pre         0.26012    0.07436   3.498 0.000536 ***
## x.CO          7.70517   10.25587   0.751 0.453035
## x.NO          0.05886    0.22250   0.265 0.791524
## x.NO2         0.58549    0.22439   2.609 0.009506 **
## x.O3          0.25979    0.07698   3.375 0.000831 ***
## x.SO2         1.27292    0.74207   1.715 0.087259 .
## x.RH         -0.18142    0.08397  -2.161 0.031472 *
## x.WS         -1.54797    2.31876  -0.668 0.504883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.697 on 316 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.5742, Adjusted R-squared:  0.5635
## F-statistic: 53.27 on 8 and 316 DF,  p-value: < 2.2e-16
```

我們把不顯著的變數刪掉，得到下列結果

```
##
## Call:
## lm(formula = y ~ x.pre + x.NO2 + x.O3 + x.CO + x.RH)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.040  -5.315  -1.153   4.713  42.751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.49089    6.14876   1.544 0.123686
## x.pre         0.30519    0.06451   4.731 3.36e-06 ***
## x.NO2         0.46569    0.18435   2.526 0.012016 *
```

```
## x.O3      0.25080    0.06911    3.629 0.000331 ***
## x.CO      16.11855    7.85327    2.052 0.040938 *
## x.RH      -0.20765    0.07288   -2.849 0.004668 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.785 on 320 degrees of freedom
## (4 observations deleted due to missingness)
## Multiple R-squared:  0.5619, Adjusted R-squared:  0.5551
## F-statistic: 82.1 on 5 and 320 DF, p-value: < 2.2e-16
```

分析結果 R-Square 為 0.5619416, adjusted R-square 為 0.5550969, 顯著項為前一天 PM2.5、 NO_2 、 O_3 、 CO 、相對濕度迴歸模型為

$$y = 9.4908928 + 0.3051878x_1 + 0.4656896x_2 + 0.2508018x_3 + 16.1185522x_4 - 0.2076531x_5 + \epsilon$$

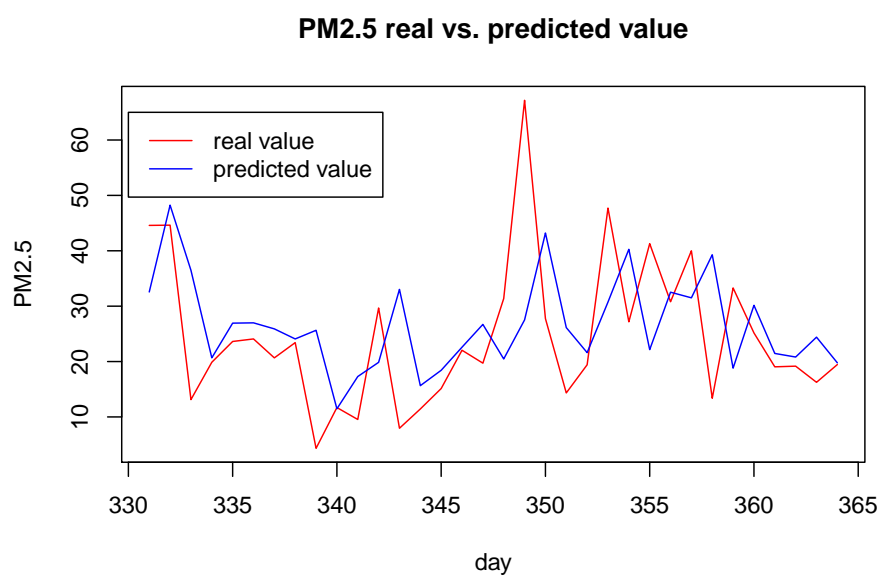
step 3: 根據所分析資料，預測 PM2.5 值

我們把 2015 年第 331 天以後變數帶入求得的回歸模型做計算，求得隔天的 PM2.5 預測值，並與實際值做比較，下表是 2015 年第 331 天以後的預測值和實際值以及他們的差值

預測值	實際值	預測值-理論值
32.56366	44.583333	-12.0196691
48.23960	44.625000	3.6145984
36.48012	13.125000	23.3551152
20.66981	19.869565	0.8002430
26.93401	23.625000	3.3090055
26.98802	24.083333	2.9046861
25.89856	20.666667	5.2318932
24.08206	23.416667	0.6653966
25.64592	4.333333	21.3125837
11.46882	11.695652	-0.2268346
17.29384	9.541667	7.7521761
19.85826	29.666667	-9.8084086

預測值	實際值	預測值-理論值
33.02714	7.958333	25.0688061
15.64710	11.434783	4.2123134
18.42045	15.125000	3.2954534
22.59529	22.041667	0.5536243
26.70733	19.708333	6.9990009
20.46798	31.375000	-10.9070197
27.52414	67.166667	-39.6425313
43.20691	27.750000	15.4569093
26.13006	14.333333	11.7967235
21.61063	19.416667	2.1939630
30.74114	47.708333	-16.9671916
40.26396	27.173913	13.0900516
22.15066	41.291667	-19.1410083
32.52219	30.791667	1.7305253
31.50597	40.000000	-8.4940307
39.28278	13.375000	25.9077798
18.80636	33.291667	-14.4853029
30.16369	25.166667	4.9970192
21.46269	19.045455	2.4172357
20.81606	19.166667	1.6493958
24.40620	16.250000	8.1562050
19.76571	19.458333	0.3073739

其中，差值 <10 的項數占了 62.8571429%，另外，附圖是 PM2.5 理論值和實際值的折線圖



結論: 預測結果 R Square=0.57, 還算 OK