

周末外出是笨蛋？---Two Sample Test 應數博 96 林良靖

有人這麼認為，周末的時候待在家裡比較好，因為周末的空氣品質比平日還差，這句話是對的嗎？類似的問題很多，這兩種情況到底有沒有差異，或是說這兩群樣本的平均數是否相同，所用到的手法就是群組間樣本的檢定，這篇文章先僅以兩個群組來做說明。我們將以 2007 年 11 月份，高雄市前金區的空氣品質來做分析，資料來源為行政院環保署所測量到有害氣體的濃度比率。

我們的空氣質素，是受道路上的車輛和工業所排放的污染物所影響的。車輛所排放的廢氣含懸浮粒子、碳氫化合物和氮氧化物；而工業也會排放污染物如二氧化硫等。我們吸入這些含污染物的空氣，會感到不適並對健康造成長期不良的影響。當這些污染在空氣中的數量超過一定環境保護處的特定標準時，我們便稱現時的空氣正受到污染，而『空氣污染』的名稱便是這樣來的了。對人體有一定程度傷害的有害氣體，列表如下：

① 二氧化硫(SO₂)：二氧化硫是無色、容易產生作用的氣體，主要來源是燃燒礦物燃料的發電廠及工業鍋爐。另一個來源是汽車排放廢氣。

② 臭氧(O₃)：臭氧並非由市區或工業區某些特別工序所直接導致排放於大空中，因此只能視為一種次生污染物。它的生成和濃度與初生污染物及紫外線有直接關係。

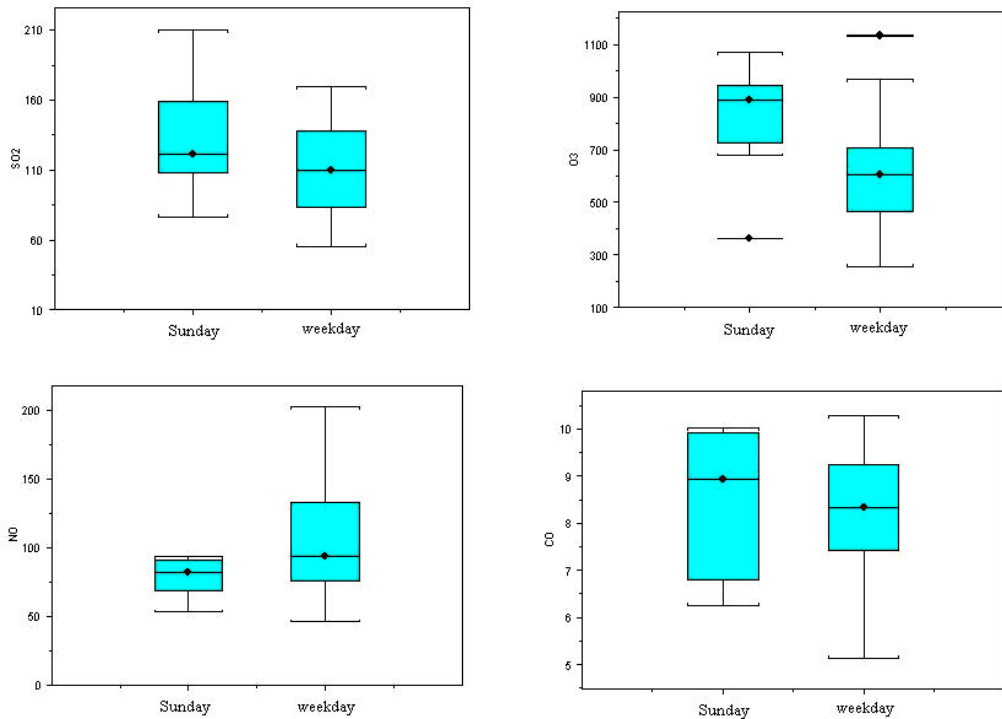
③ 一氧化碳(CO)：一氧化碳是無色、無嗅、無味的氣體，是在不完全燃燒碳的情況下產生的副產品。工業程序也會增加一氧化碳污染濃度，但在大部分的市區，例如在香港，一氧化碳的主要來源則是汽車。

④ 氮氧化物(NO_x)：當燃燒序以高溫進行時，例如發電廠及汽車引擎操作時，這等氣體便會形成，排放入大氣中。一氧化氮(NO)是燃燒過程中所生的主要的 NO_x，會轉化為二氧化氮。而二氧化氮是市區霞氣或光化學霧的主要成分。

接著再從行政院環保署所取得的資料做分析。由於資料量過大，在此不將其列出，僅做簡單的說明；所得的資料為 SO₂、O₃、NO 與 CO 在空氣中的濃度比

率，每個小時測量一次，單位為 ppb(parts per billion, 即十億分之一)。其中有一些數據是被儀器檢核為無效值，此情況將利用前後兩筆的平均值來取代。

此外，我們只取一般活動時間(早上八點到晚上六點)的測量值，分成周末與非周末兩群，分析周末有害氣體的比率是否較高。首先觀察盒鬚圖(Box-Plot)如下：



比較上面的四個圖，發現除了臭氧(O₃)在兩群間有明顯的差距外，其餘有害氣體無明顯的差距；然還是需要用檢定的方法來做確認。檢定兩群樣本平均數有否差異，t 檢定是一個好用的方法； μ_1 代表周末時的濃度、 μ_2 代表非周末時的濃度。

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2$$

其統計量為：

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n + 1/m}} \sim t_{m+n-2}$$

而 m 、 n 分別代表兩組樣本的個數； $S_p^2 = \frac{(n-1)\hat{S}_X^2 + (m-1)\hat{S}_Y^2}{m+n-2}$ 。必須注意此檢定的

的必要條件，即要求兩組樣本皆需服從常態分佈。在此我們將使用檢定力較高的

Shapiro-Wilk test，來檢定所有的數據是否來自常態分佈：

	周末	非周末
SO2	0.729	0.654
O3	0.242	0.089
NO	0.383	0.192
CO	0.083	0.758

表格的值皆為機率值(p-value)。由上表可知，皆接受虛無假設的論述，即所有的群組皆服從常態分佈。接著用 t 檢定與另一種小樣本常用的檢定 Wilcoxon rank sum test (詳細用法請參考文獻一)來檢定『周末有害氣體的濃度不同於平日』這句話是否為真。對各種氣體檢定後整理機率值 (p-value) 如下：

	t test		Wilcoxon
SO2	0.932	SO2	0.983
O3	0.040	O3	0.039
NO	0.133	NO	0.142
CO	0.742	CO	0.389
O3*	0.022	O3*	0.019

上表中最後一列，是針對臭氧的資料拒絕了虛無假設，即認為這兩群平均數是不相同的；故我們進一步做檢定為 $H_0: \mu_1 - \mu_2 < 0$ versus $H_1: \mu_1 - \mu_2 > 0$ 所得的 p-value 值；由上表的結果可以有下列結論：高雄市前金區的空氣有害氣體，除臭氧於周末時的濃度明顯高於非周末時的濃度；其餘的有害氣體則無明顯的證據證明周末時的濃度較高。而前面亦有敘述，臭氧為次生污染物，主要由紫外線產生，亦有可能只因為周末的溫度較高才導致這個結果。因此，周末時還是多向戶外走走，紓解平時所累積的壓力，以保持最佳的身心狀況。

參考文獻

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis, second edition*.
Wadsworth Publishing Company, Belmont, California.

行政院環保署 <http://taqm.epa.gov.tw/emc/default.aspx?mod=PsiAreaHourly>