

以抽樣法估計攤販數量－以桃園區為例

徐浩然

桃園區公所會計室

(2017/12/15)

在桃園區(以下通稱本區)，統計攤販數量乃按月辦理且行之有年之例行性工作。惟本區地屬都會區且幅員廣大(佔地 34.8047 平方公里)，採地毯式普查頗不符成本。於本文中，我們採用生態學家常用來估計生物族群數量之「區塊法」(Quadrat Sampling)，對本區的攤販數量進行估計。我們先以 Google Earth 對本區範圍進行區塊化，並將每個區塊依序編號。緊接著隨機抽取固定數量的區塊以利後續訪查。訪查前將該區塊的邊界資訊匯入 Google Map 以便訪查時可精準定位。最後利用訪查所得資料對攤販數量及其變異數進行估計。本文希望透過現有免費雲端工具之充分利用、以及採用具統計理論基礎之抽樣方法，使估計攤販數量的工作更加系統化與科學化。

前言

根據本公所會計室過去資料顯示，統計攤販數量的工作早於民國 89 年起即開始，並且須按月填報公務統計報表「桃園市桃園區攤販數量及取締績效」(表號 10959-04-51-3)。儘管該統計工作已行之有年，但據了解，目前並無任何明文規範統計的進行方式。對於地屬鄉間的區域，也許直接調查熱鬧街區即可獲得充分的統計數量；但對於地屬都會的本區來說，就不適合採用。本文將以「區塊法」¹ (Quadrat Sampling) 為抽樣的理論基礎，搭配 Google Earth² 為抽樣工具以及 Google Map³ 為實地訪查工具，最後以區塊法的公式計算攤販總數與其變異數的估計量。

區塊法

假設總數為 M 的攤販隨機分布於面積為 A 的本區內，且令 A 可切割成 N 個面積為 a 的小方塊，亦即 $A = Na$ 。從中隨機選取 n 個小方塊，令 m_i 為第 i 個方塊的攤販數，且

$$M = \sum_{i=1}^N m_i。$$

令單位面積的攤販數量 $\lambda = \frac{M}{A}$ ，故 $M = \lambda A$ ，其估計量為

$$\hat{M} = \hat{\lambda} A \quad (\text{方程 1})，$$

其中

$$\hat{\lambda} = \frac{\bar{m}}{a} \quad (\text{方程 2})$$

並且

$$\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i \quad (\text{方程 3})，$$

變異數的估計量為

$$\hat{V}(\hat{M}) = A^2 \hat{V}(\hat{\lambda}) = A^2 \left(\frac{\hat{\lambda}}{an} \right) \quad (\text{方程 4})。$$

地圖區塊化

先至政府資料開放平台下載 GML 格式的「鄉鎮市區界線(TWD97 經緯度)」資料集，解壓縮後以 wordpad 軟體開啟名為

「TOWN_MOI_1060503.gml」的檔案。在該檔內以「桃園區」為關鍵字進行搜尋，可找到本區的邊境經緯度資料。將其複製並貼入至 excel 檔，做好整理後，再匯入 Google Earth 軟體使本區邊境顯示出來。利用該軟體縮放功能將經緯線交織成的方塊縮放至適當大小(圖 1)，使方格能夠涵蓋區大部分範圍。接下來的重點為方塊的選擇：完全被包含在本區範圍內的方塊為必選；至於被本區邊界線貫

¹ Richard L. Scheaffer 等原著。鄭天澤譯。2014。抽樣調查。初版，349-353。新北市：高立。

² <https://www.google.com.tw/intl/zh-TW/earth/>

³ <https://www.google.com.tw/maps/>

圖例

- 方塊編號
- 抽樣方塊



4 km

N25°01'48"



1
圖

Google Earth

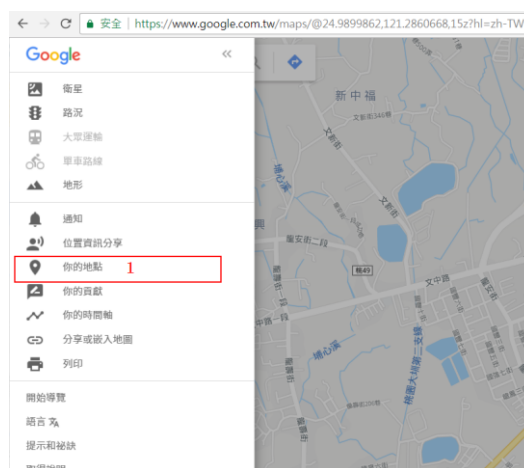
Image © 2017 DigitalGlobe
Image © 2017 CNES / Airbus

抽樣

(圖 2)

訪査

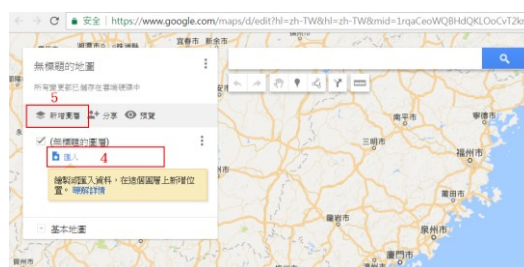
接下來需將抽樣方塊的 kmz 檔匯入至 Google Map。首先於 Google Map 進入[你的地點](圖 3 步驟 1)



(圖 3)

(圖 4)

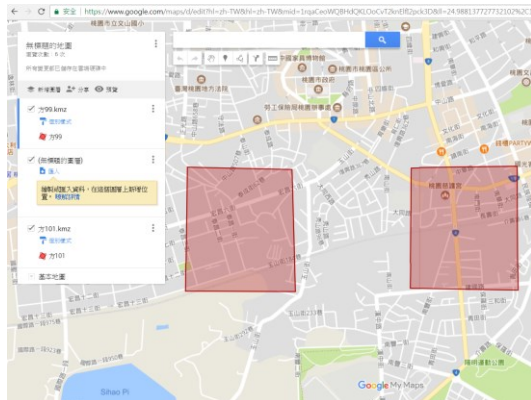
經由(圖 5 步驟 4)匯入第一個抽樣方塊的 kmz 檔，其餘的 kmz 檔可透過[新增圖層]方式匯入(圖 5 步驟 5)



(圖 5)

最後在 Google Map 上會出現紅色方框(圖 6)，

訪查員於事前在手機安裝好 Google Map app，就可以在紅框範圍內進行實地訪查。



(圖 6)

各抽樣方塊的訪查結果如(表 1)所示。

方塊編號	67	95	99	101	122
攤販數量	0	0	0	1	1

(表 1)

估計

由(方程 3):

$$\begin{aligned}\bar{m} &= \frac{1}{n} \sum_{i=1}^n m_i \\ &= \frac{1}{5} \sum_{i=1}^5 m_i \\ &= \frac{1}{5} (0 + 0 + 0 + 1 + 1) = 0.4(\text{個})\end{aligned}$$

由(方程 1)與(方程 2):

$$\begin{aligned}\hat{M} &= \hat{\lambda}A \\ &= \frac{\bar{m}}{a} \times A \\ &= \frac{0.4}{0.27} \times 35.91 \cong 53(\text{個})\end{aligned}$$

再由方程 4:

$$\hat{V}(\hat{M}) = A^2 \left(\frac{\hat{\lambda}}{an} \right)$$

$$\begin{aligned}&= 35.91^2 \times \left(\frac{0.4}{0.27 \times 5} \right) \\ &= 1415.12(\text{個}^2)\end{aligned}$$

$$\sqrt{\hat{V}(\hat{M})} = \sqrt{1415.12} \cong 38(\text{個})$$

故在一定的信心水準之下，本區的攤販數量介於 $53 \pm 2 \times 38$ (或-23 至 129) 之間，此誤差區間明顯過大，甚至跨越了負值區域。為改善此一情形，由(方程 4)可知，增加抽樣方塊數目 n 或增加抽樣方塊面積 a ，可以減少誤差，惟依照我們實地訪查情形，騎 100c.c. 機車訪查單一面積為 0.27 平方公里的抽樣方塊，耗時約 30 至 40 分鐘，此為需考量的現實面問題。

結語

在本文中，我們視攤販隨機散布於二維固定有限範圍之內，然後根據「區塊法」的抽樣理論基礎，提出以 Google Earth 與 Google Map 為訪查工具之攤販數量估計法。我們認為，對於幅員廣大且人口稠密的地區，採用傳統方法，諸如普查法或根據經驗直接調查攤販易聚集處，皆各有其不足之處：(1) 前者須花費大量人力與時間、(2) 後者則無法適應攤販的流動特性。因此我們相信，採用抽樣法可解決前述之所有問題。儘管目前所得到的估計結果因抽樣誤差過大而不如預期，但該誤差正好提供了重要資訊，指引我們未來提升估計精準度的努力方向。