

第 7 章：在 GIS 裡統計資料的空間分析

2
8
1
3
7

7.1 介紹

視覺化可以幫助我們了解空間的情形，但是缺點是需要更多的調查來描述或解釋這個部分。量化的資料，這需要研究者去執行某些統計分析。在使用 GIS 時，分析是應該要被使用的，但是不只是使用傳統的統計技術，傳統的統計技術只集中在屬性資料；其實應該結合空間的元素。這兩者結合稱做「空間分析」或是「資料分析」(GDA)。

使用以 GIS 為基礎，分析空間資料時，需要小心謹慎；而 GIS 在人文地理方面的研究，是可以學習參考的。

在早期，GIS 為基礎的做法，聲稱 GIS 提供了一個有凝聚力的和科學的框架，可以重新團結地理作為一門學科(see, for example, Openshaw 1991)。意料之中的，這導致反彈，特別是著重於定量和科學方法的研究的人。對此，有人覺得說，這是回到很差的實證主義。也有人覺得說，GIS 缺乏一個強而有力的認識，以及可以處理倫理、經濟或政治的問題(see, in particular, Pickles 1995)。

隨著時間的過去，一些比較極端的要求與批評，都已趨緩或解決。特別的，與單純提供質性、實證的系統比起來，GIS 明顯得可以提供更多東西。另外在第八章會討論，在定性與定量的研究，GIS 也可以發揮重要的作用。

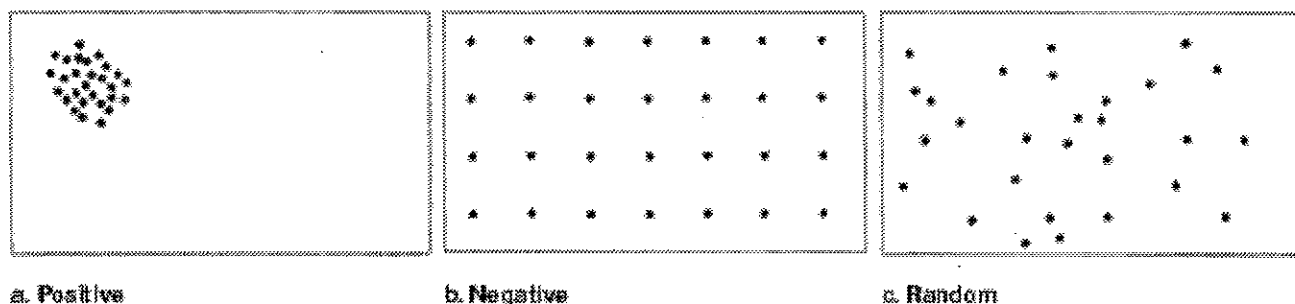
對研究者而言，有很多好用的統計功能，像是 GIS toolbox。許多套裝軟體要求有分析的功能，但這通常與 overlay、buffering 和其他空間操作有關。這是在第四章所描述的。特別的是，GIS 支援研究者，以確定需要的是哪類型的技術，還有跟它相關的技術，並制訂方法，表現分析。目前我們需要的，是我們的數據需要甚麼樣的資料。使用填鴨式的方法或用供應商提供的解決方案可能並不適當，這會導致繁瑣的計算。我們希望能讓數據有更適當的分析方法。

7.2 是什麼東西使得空間參考資料變得特別？

空間參考資料為想要執行統計學分的研究者提供了優缺點並存的奇異特性；根本上的優勢即為允許如「這是在哪裡發生的」、「這現象是如何在研究區中變化的」、「此地點所發生的事件如何影響週圍地區」與「在單一變數比例較高的特定地點，是否有其他比例較高的變數」等類型的問題探究，傳統的統計學如相關分析或回歸分析傾向於產生一個概括的統計量，此統計量即為將資料設置中特定兩個(或以上)的變數加以數量化的結果，舉例來說，如 X 與 Y 的相關係數為 0.8；這個被稱為全球或全圖分析的方法從 GIS 的觀點來看並不受到支持，因為其僅提供了整個研究區平均關係的摘要，而忽略了空間的影響；事實上，任何關係均會越過空間而變化，當以空間參考資料進行分析時，應使用能檢查整張地圖中平均關係，以及這些關係如何變化的技巧來實行，這被稱做地方性分析(Fotheringham 1997)，在 7.3 會有實行這些分析的案例，現在我們可以藉著使用明確融合資料空間的技巧來達到更細膩的分析解讀，類似傳統分析一般地忽略空間的影響，將限制我們能從資料中擷取的資訊。

空間參考資料附帶了四個主要的缺點，分別是：資料品質與錯誤擴散、空間的自動交互作用、可調整區域單元問題(MAUP)與區群謬誤；關於空間參考資料的資料品質已在第四章中討論過，

但可能須考慮到所使用方法的類型。空間資料的類型經常由什麼樣的空間分析技術是最適當來決定的，因為不同的技術必須使用一些利用點、線和多角形資料集的特徵。它是有時根據他們矩心所在而將多角形資料視為點來分析，或將點作為多角形來使用 Thiessen 多角形(參見章節4)。



空間自相關係數恰當的點模式的類型

點模式分析是企圖去確認是否是隨機的分佈點，或是否集中一地或者集群（正面空間的相關係數），或是均勻分佈的（消極空間的相關係數），如圖7.1所示。

以最簡單的方式來做，這被稱為「樣分析」(*quadrat analysis*)，即研究區域被細分成是規則柵格正方形，且並且事件的數量會在各個正方形中來計數。

不過這不是特別理想的，因為結果往往會太依賴柵格正方形的大小規模和排列方式。更好的技術是側重於每個點和其最近的鄰近點，或對內核心之間距離的估計，在地圖上每個位置從而有效地產生一個移動平均線的密度點。

全球性的統計摘要，或者從這些方法中任意簡單地說，數據聯網資料成群是均勻分佈的抑或是隨機分佈。若能使用當地技術，就更好、更容易地能夠提供重要摘要，例如「數據群聚在這個位置，而非是隨機分佈在這裡。」

純點模式分析只不過是關注在分析數據空間的組成部分和忽略其屬性。對於多邊形數據是不可能去完成的，那是因為多邊形數據的排列通常是多變的，如以人口普查數據為根據上的行政單位。

對於這些是有必要策劃鄰近的度量以使我們能夠去量化每個多邊形已就其鄰近點的影響。這些既可以是基於在矩心之間的距離，或基於多邊形的界限內某一度量關係之間來完成。

他們也可能是二進制多邊形 *i*，或者是具有影響力的多邊形 *j*，或者都不是，也或者是可能有某種程度量化對強度的關係。簡單二進制接近的度量，包括兩個多邊形是否共享邊界，或者他們的矩心是否在範圍內展現彼此間的一段距離。更複雜的方式包括比例的基礎上，共同邊界的長度在多邊形間相比，其總周長或距離呈現遞減模式，當距離在兩者矩心之間增加時，藉以影響另一個已傾斜的多邊形距離。

測量距離尺寸，可用於分析多邊形或點的數據資料。一個單純形狀的多邊形使用可分析距離和屬性的 *Geary's coefficient* 或 *Moran's coefficient* 來做空間相關的測試。原則上其付予全球性概要性探討，不過呢像 *Geary's Gi* 的科技讓我們可以針對這些問題處理區域的分析。更先進的空間分析樣式則結合了包含克里格法 (*kriging*) 和地理加權回歸 (*geographically*

將這兩個分層區塊疊合則能將倫敦再細分成四個類型的區塊：有缺乏排水和受污染的水源供應風險的區塊、只有缺乏排水風險的區塊、只有受污染水源供應風險的區塊以及兩種風險因素都沒有的區塊。然後他們計算每個類型區塊中因霍亂死亡的比例。即使簡單的圖表化這些資料，都顯示出一個清楚的模式：有兩種風險因素的區域之死亡率幾乎是大都會區平均的兩倍高。只有排水缺陷的區域之死亡率也是高於平均值，只有水源污染的區域其死亡率約等於平均值，而兩種風險因素都沒有的區域之死亡率則低於平均值。傳統上，統計分析會使用變異數分析來驗證這個模式。

Gregory (2000) 完成對人口調查和重要登記資料分析的來顯示怎麼運用空間和時間資料的組成來增加關於屬性的資訊。他透過 1881 年到 1931 年的 50 年期間遷移網來做這些。人口調查提供人口資料依每五年年齡帶和性別作分割。*The Registrar General's Decennial Supplements* (註冊總署) 提供年齡層和性別的死亡數量。在理論上這提供足夠的資訊來計算在 10 年間的年齡層和性別的遷移。舉例來說，從 15 到 24 歲，這十年間的婦女的數量減去是 5 到 14 歲的婦女的數量，然後減去死亡的數量，應該可以算出這一隊人的遷移率。界限變動使這個演算具有高度錯誤傾向，因為任何一個由界限變動造成人口變動將被視為淨遷移。一個單獨地區的區域，年齡和性別，具體淨遷移率的所有資料的插值法可能長期在區級中被計算。因此，這分析接受非常基本的人口統計學的資料並使用它提供常的長時間序列，比新資料的空間資料和屬性資料更有可能從手工方法被創造出來。

Martin(1996) 比較 1981 年和 1991 年兩次人口調查。他從他感興趣的 200m 為邊的方形映像點網格的兩次人口調查中內插了資料。這樣他比 Gregory 更能在二次人口調查之間與比較更多空間資料的細節，但無法包括更加早期的人口調查因為他們沒有提供充足的空間細節來允許準確插值法。

7.5 結論

雖然一個有意義的數量進步已經在確定研究統計資料的定量分析議程過程中取得，在這功能性成為一個GIS工具箱的標準部分之前，仍然有一些方法可使用，甚至更進一步在它成為一個歷史分析的普通部分之前。有大量可能性為空間引用的數據的統計分析，但是必須記得數據的限制因素。這意味研究人員必須選擇不倚賴獨立的偶然性的假定的技術，並且必須記住對他們的數據的限制，尤其是它的特性：錯誤宣傳引起的潛在的問題；可修改表面的單位的可能影響；以及由於生態謬見的錯誤推論的可能性。

史學家中使用GIS變得越來越複雜，增加使用空間統計方法這是被期望的。事實上，迄今為止提供機會對研究人員想出非常適於他們自己的數據的技術是不容易結合於GIS軟體中，雖然這些必須被在對來源數據的限制內設計出。數據的空間本性的影響一定在這裡被記住，尤其，不必要的聚集應該被避免不僅透過空間，而且整個屬性和時間。這樣技術不應該對適合整個研究地區只提供一個單摘要統計而是透過地圖應該顯示現象變化怎樣的影響。這種技術的結果可能以地圖為基礎而不是使用傳統的數值的摘要。