



可調整地區單元問題 (MAUP) 對分析人口分布的影響

人口是從事各項社會與經濟研究中重要的一環，人口與社會經濟發展資料可利用地理資訊系統整合至空間單元，探究其中之關聯，以作為資源分配與區域規劃的依據。而進行空間資料彙整時，其成果與展現，將受到可調整地區單元問題而有不同的判讀與分析，因此，本研究使用核密度估計法、區域性空間自相關、皮爾森相關分析進行，探討最小統計區與各級發布區的資料是否能反應真實的地理現象。

張國楨、歐采瑀、王翊芬、洪怡婷、徐品翰（國立臺灣師範大學地理學系教授、群琬地理資訊顧問股份有限公司專案經理、國立臺灣師範大學地理學系學生、學生、學生）

壹、前言

隨著社會經濟環境快速變遷，居民對於社區意識逐漸提升，社區發展、公共建設成爲大眾密切關心的議題。然而在資源有限的情況下，如何有效的將資源分配運用以獲得最大效益，是社會發展與建設時所需考量的部分，其包括了人口與社會經濟的整合與應用。在進行社會經濟與人口資料分析與應用時，必須考慮以下三點：

一、可調整地區單元問題 (Modifiable Area Unit Problem, 簡稱 MAUP)

資料搜集時空間單元的差異所引發的空間分布變異。

二、空間相依性 (spatial dependence)

係指空間中存在的現象並非獨立或隨機發生，相鄰的空間單元會產生相同現象群聚的

空間關聯情形，是地理分布現象和空間過程的本質特徵。

三、空間異質性 (spatial heterogeneity)

係指空間中存在的現象受到個人愛好與態度、環境等條件差異的影響，會隨著地理位置不同而產生變數間關係的改變。

進行社會經濟、國土計畫等研究時，若以原始資料進行探討，將最能貼近實際現

況，降低分析錯誤的機率，然而由於近年民衆對於隱私需求日漸提升，個別資料取得困難，導致相關研究環境惡化。有鑑於此，臺灣跟進先進國家，陸續建置統計區系統作為各類別資料蒐集、彙總及發布之基本單元，並在保護隱私前提下，提供小區域或特殊區域統計以滿足各界需求，如美國的 Tiger 系統 (Topologically Integrated Geographic Encoding and Referencing)、加拿大的 SAC 系統 (Statistical area classification) 等。臺灣的統計區分類系統如圖 1 所示，包括最小統計區 (Basic Statistical Area) 為資料彙總的最小空間單元，以及資料發布使用的各階層發布區 (Dissemination Area)，由最小統計區、一級發布區、二級發布區、鄉鎮市區至縣市等，其中，最小統計區所含的個體資料具均質之特質，可應用於多種層面，如工廠區位選址、人口變遷等。

統計區的建立能彌補以行政區作為資料彙整、統計單元常發生之分區單元不夠細緻及行政單元於時間上的重

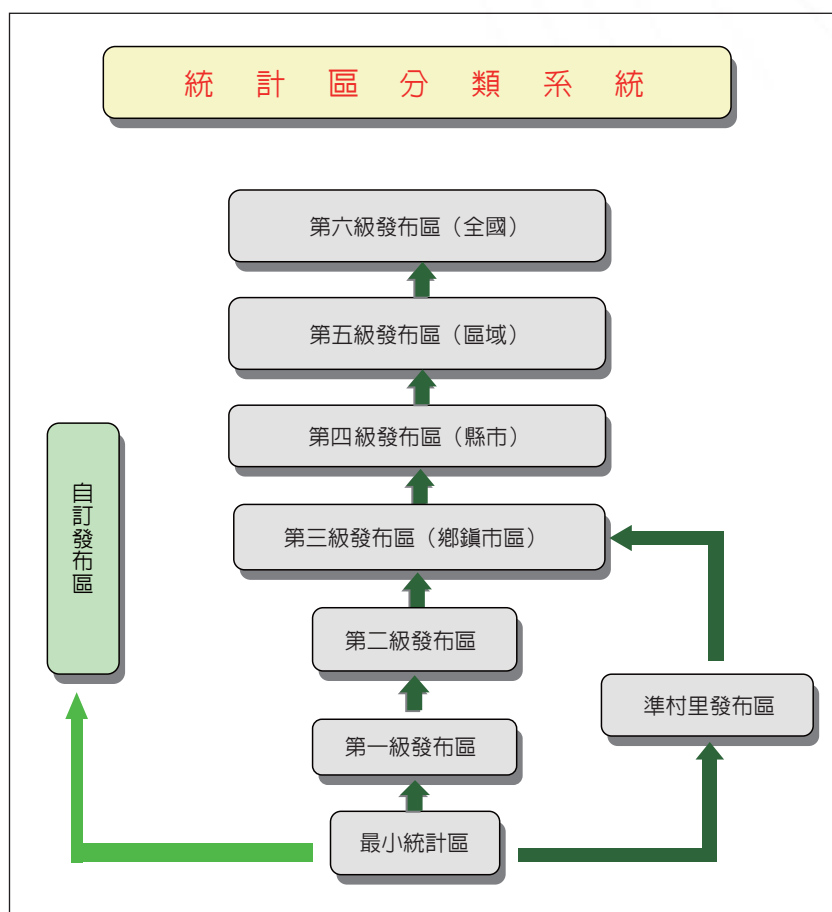
組與邊界變遷等問題，造成資料實際的空間分布型態不一致，難以合理比較，影響後續資料應用價值。目前國土資訊系統已積極使用最小統計區作為資料的統計單位，然而為探討最小統計區的統計資料與真實地理現象分布是否具有差異性。本研究以臺北市作為研究範圍，探討在 MAUP

下資料空間分布之異同性，並進一步探討最小統計區的統計資料與真實地理現象分布之差異性。

貳、研究目的

在都市管理和規劃之中，空間單元的大小是一項重要的變因，然而為方便行政管理為考量，以行政區作為資料彙整、

圖 1 統計區分類系統



資料來源：內政部統計處。

論述 》 統計 · 調查



統計之單元會產生分區單元不夠細緻及行政單元於時間上的重組或邊界變遷等問題，造成資料與實際的空間分布型態不一致，形成 MAUP 的謬誤，進而影響後續資料的效用與價值。本研究希望就空間相依性與異質性進行初步探討 MAUP 對人口分布的影響。因此，本研究目的有二：

- 一、探討可調整地區單元問題對人口分布的影響。
- 二、探討最小統計區的統計資料與真實地理現象分布之差異性。

參、研究流程與方法

一、研究流程

本研究流程如圖 2 所示，探討之空間單元由最小統計區至三級發布區（村里單元）。首先將人口資料整合於各地區單元圖層中，分析在不同地區單元下，人口分布之差異性。為探討人口分布之差異性，將使用核密度估計法，計算人口於各地區單元空間中之密度，並透過核密度分析法中搜尋半徑的調整，探討 MAUP 所產

生的謬誤。而本次研究透過皮爾森相關分析，探討人口核密度估計與搜尋半徑之相關性，並以區域性空間自相關（Getis-Ord G_i^* ）之 G_i^* 值轉換為 Z 值進行顯著性檢定，以百分之 95 信心水準之下，Z 值 >1.96 為基準，分析臺北市人口在空間中熱點分布的顯著性，探討不同空間單元下，人口集中區域分布的差異性，並探討相鄰層級之地區單元是否能相互取代。

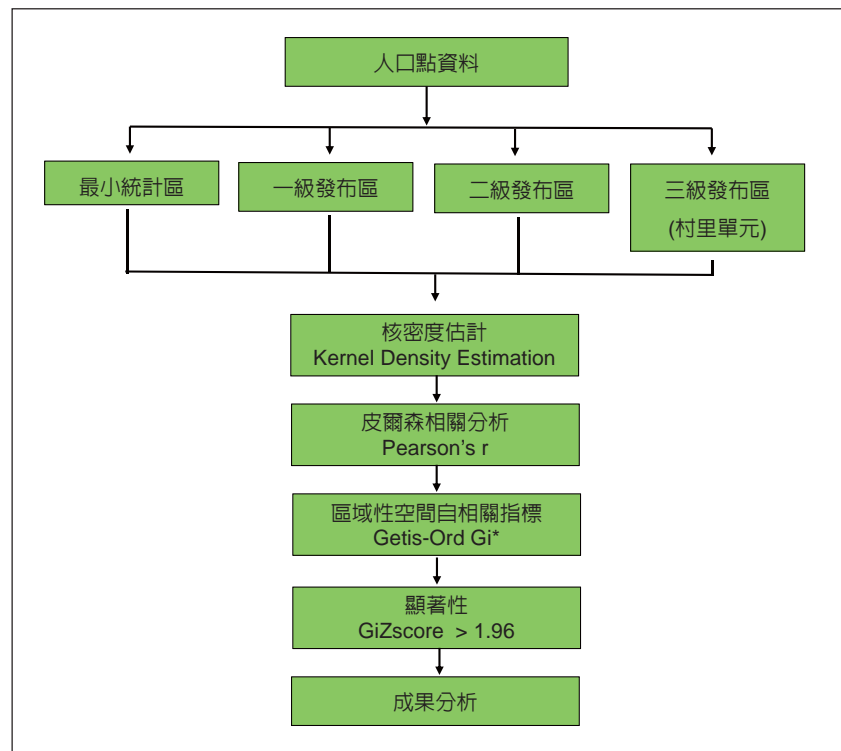
二、研究方法

（一）核密度估計 Kernel

Density Estimation

本次研究以核密度估計臺北市人口在空間中密度分布的特性。此估計法建立於均勻的網格，以網格的中心點為中心，設定搜尋半徑以作為估算範圍，計算特徵值在鄰近區域的密度，當搜尋半徑愈大時，其呈現資料的概括化程度愈高，會降低空

圖 2 研究流程



資料來源：作者自行整理。

間單元的區隔問題，使其呈現較為平滑的空間樣貌。

公式如下 (1)：

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K \left\{ \frac{1}{h} (x - X_i) \right\} \quad (1)$$

$\hat{f}(x)$: 表示核密度函數的密度值

h : 表示搜尋半徑

d : 空間次元

n : 表示點資料的數量

x : 表示座標向量

X_i : 在第 i 個位置的座標向量

本研究透過核密度分析，將不連續的點資料以連續的曲面呈現人口聚集與分布之區域，當核密度值愈大，代表聚集的程度越高。

(二) 區域性空間自相關指標

Getis-Ord G_i^*

為探討空間資料在不同統計單元呈現的差異與找尋人口分布之熱點。本研究使用區域性空間自相關 Getis-Ord G_i^* 統計方法。此研究方法對每一個統計單元會賦予一數值，代表目標空間單元與鄰近空間單元其屬性特徵在空間中的聚集程度。若空間單元中一群高 G_i^* 值表示高值集中，即為熱點 (Hot Spot)，空間單元中一群低

G_i^* 值表示低值集中，即為冷點 (Cold Spot)， G_i^* 值接近 0 代表特徵附近沒有高值或低值集中的現象。於本次研究中將進一步對 G_i^* 值做顯著性分析，以 $Z > 1.96$ 為基準，檢視各統計單元於不同搜尋半徑下，熱點呈現之情形。

公式如下 (2)：

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j} \quad (2)$$

$w_{ij}(d)$: 距離為 d 的範圍內相鄰單位間的權重大小

x_j : 空間單元中之屬性數值

本研究中，熱點代表統計值有高度集中的現象，表示此區域人口較為集中，而冷點代表統計值為高度分散的現象，表示此冷點區域人口較為稀疏。

(三) 皮爾森相關分析 Pearson's r

為探討可調整地區單元問題，本研究以皮爾森相關係數分析人口分布與搜尋半徑的關聯性。皮爾森相關係數是用來表示兩個變項間線性相關。相關係數的值 (r 值) 介於 -1 到 +1 之間，係數的絕對值越大，

代表相關程度越高，係數為正代表為正相關，係數為負值，代表兩變項為負相關，若 r 值為 0，則表示變項間無關聯性。

其公式如下 (3)：

$$r = \frac{cov(x,y)}{S_x S_y} = \frac{[\sum(X-\bar{X})(Y-\bar{Y})]/N}{S_x S_y} \quad (3)$$

r : 表示皮爾森相關係數

N : 表示成對資料的對數

S_x : 表示 x 的標準差

S_y : 表示 y 的標準差

肆、研究成果

本次研究透過最小統計區、一級、二級發布區與村里單元探討人口於各空間單元下空間分布之差異性，並透過核密度估計法搜尋半徑的調整，證明可調整地區單元的問題。

一、MAUP

由核密度估計法可得知，當網格大小不變，搜尋半徑增加時，部分地區的資料會被概括化，比較戶籍點位與最小統計區之核密度，發現兩者差異極小。而為瞭解門牌人口資料以及最小統計區資料的人口核

論述》統計 · 調查

密度估計之間的分布，使用皮爾森相關係數來檢視兩種資料間的相關性，比較兩者在網格大小 150 公尺下，不同的搜尋半徑的人口密度資料的相關程度，以此了解人口的核密度估計是否與搜尋半徑的變化相關。在搜尋半徑較小的情況下，資料較為細緻，因此可以比較出較多兩項變數不同的變動方向，但仍呈現高度的正相關。當搜尋半徑漸變大時，資料不斷被均質化，使得大區域呈現的是平均值的狀態，因此相關係數越趨近於 1，表示兩者資料所呈現之人口核密度分布差異小，如附表所示。

二、各統計單元人口分布之差異性

本研究以區域性空間自相關 (Getis-Ord G_i^*) 之 G_i^* 值轉換為 Z 值進行顯著性檢定，以百分之 95 信心水準之下，Z 值 >1.96 為基準，分析臺北市人口在空間中熱點分布的顯著性，探討不同空間單元下，人口集中區域分布的差異性，以及相鄰層級之地區單元是否能相互取代。

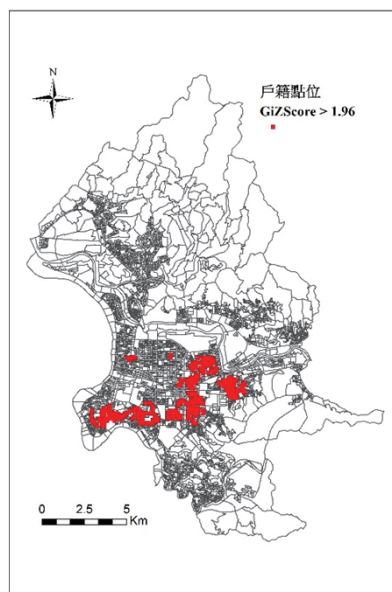
由圖 3 及圖 4 可得知當網

附表 臺北市 r^2 值

搜尋半徑	50m	100m	200m	300m	400m	500m
r^2	0.932717	0.992921	0.999237	0.999783	0.99991	0.999954

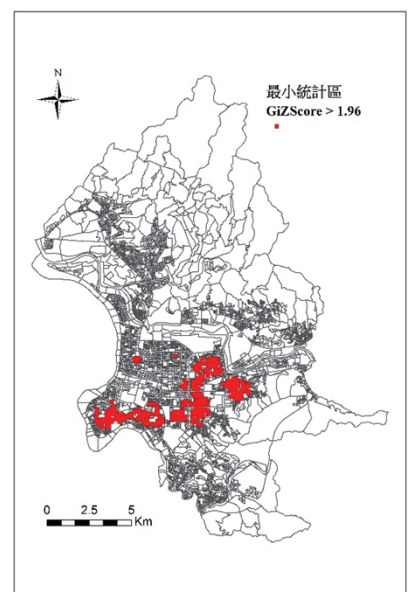
資料來源：作者自行整理。

圖 3 戶籍點位



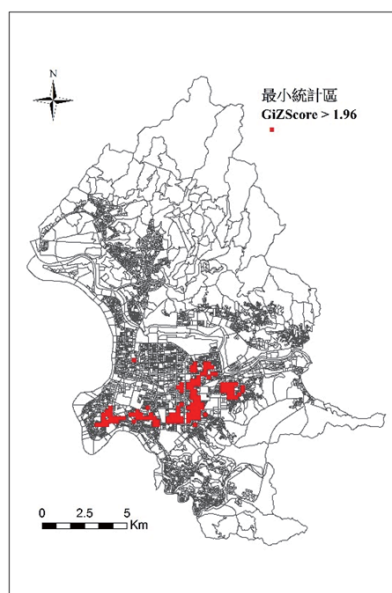
資料來源：作者自行整理。

圖 4 最小統計區



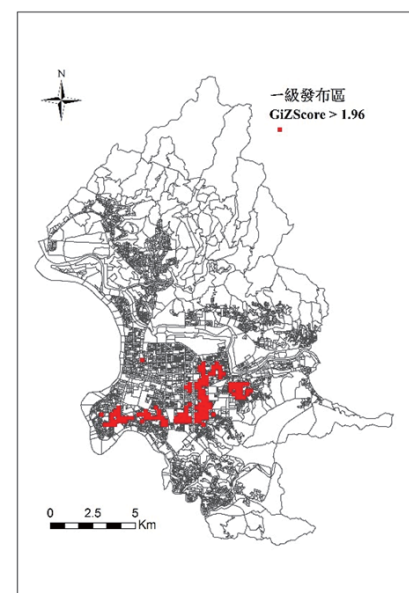
資料來源：作者自行整理。

圖 5 最小統計區



資料來源：作者自行整理。

圖 6 一級發布區



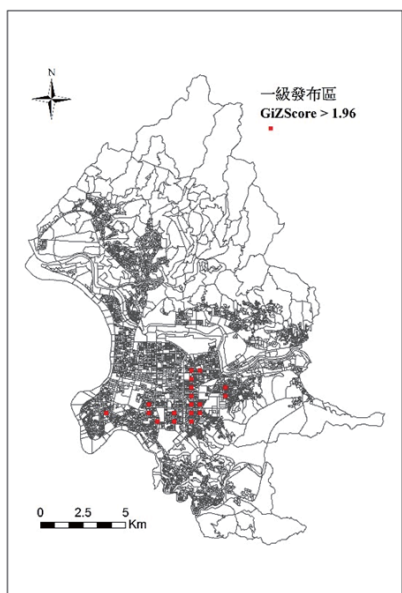
資料來源：作者自行整理。

格大小設定為 150 公尺時，戶籍資料與最小統計區的地區單元尺度之下可以看出此兩種單

元的熱點地區，差異性不高，信義區、中山區、松山區及萬華區，皆有人口熱點之區域，

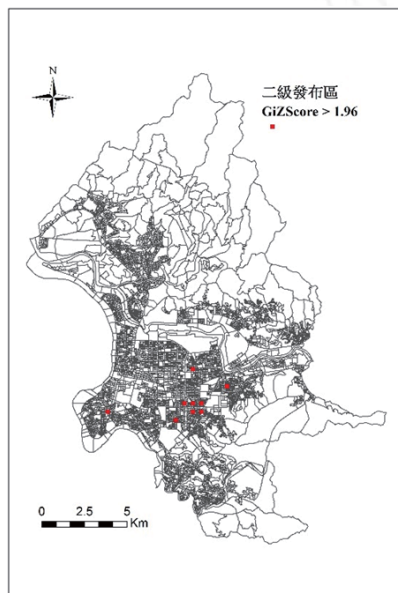
且其相互之位置近乎相同，說明最小統計區已能體現原始資料之特性。

圖 7 一級發布區



資料來源：作者自行整理。

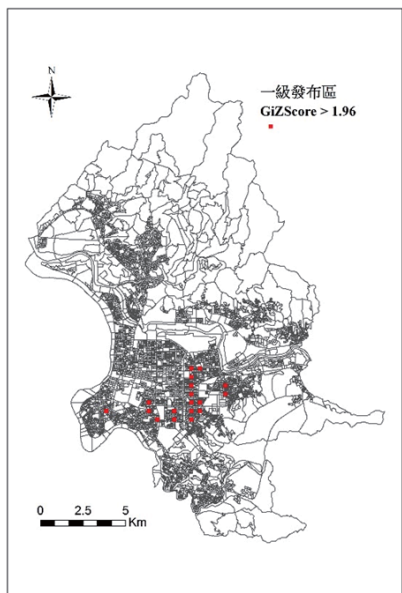
圖 8 二級發布區



資料來源：作者自行整理。

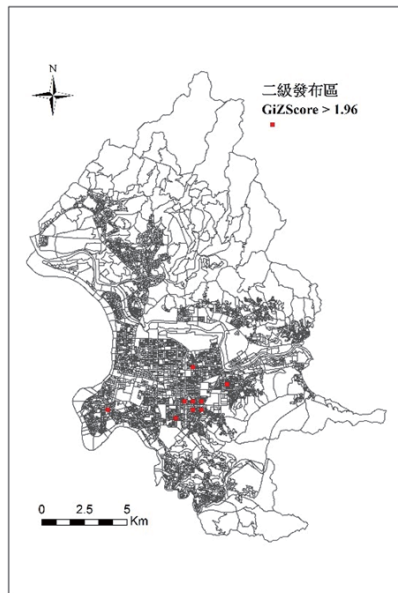
由上頁圖 5 及圖 6 可得知當網格大小設定為 200 公尺時，在一級發布區與最小統計區的地區單元尺度之下可以看出此兩種單元的熱點地區，差異性仍不高，信義區、中山區、松山區及萬華區，皆有人口熱點之區域，且其相互之位置近乎相同，說明一級發布區可體現最小統計區之資料特性。

圖 9 二級發布區



資料來源：作者自行整理。

圖 10 村里單元



資料來源：作者自行整理。

由圖 7 及圖 8 可得知當網格大小設定為 500 公尺時，在一級發布區與二級發布區的地區單元尺度之下可以看出此兩種單元的熱點地區，差異性頗高，說明二級發布區已不能體現一級發布區之資料特性。而由圖 9 及圖 10 也說明，當網格大小設定為 500 公尺時，村里單元不能體現二級發布區之資料特性。

伍、結論與建議

一、結論

(一) 可調整地區單元問題 (MAUP)

在不同空間單元下，可

論述》統計·調查



調整地區單元問題對於資料成果的展現，將會影響其成果的判讀與分析，其中尺度效應 (scale effect) 說明尺度較小的地區單元較能呈現較細緻的人口分布，而尺度較大的地區單元會使空間資料被均質化，當搜尋半徑漸變大時，資料不斷被均質化，使得大區域呈現的是平均值的狀態，無法體現空間的異質性。

(二) 各統計單元熱點分布之差異

各統計單元皆呈現不同的熱點，較小空間單元較能體現原始資料的特性，由本次研究結果可發現，各地區統計單元中，二級發布區與村里單元無法呈現細緻的人口分布，而最小統計區與一級發布區能呈現細緻的人口資料。

(三) 空間單元

本次研究以內政部統計處建置的最小統計區以及一級、二級發布區以及村里單元探討不同空間單元底下，人口分布的差異性，由結果得知越細緻的空間單元越能

充分顯現原始資料分布的特性，以最小統計區的細緻程度，能於呈現原始的資料特性的同時，保護個人資料的外洩，利於各項後續研究的進行。

(四) 最小統計區之應用

於各級空間單元，最小統計區最能保有空間統計原有之特性，並且能產生全國一致且穩定之小區域統計單元，以便進行長時間的資料比較。最小統計區可應用於國土規劃及復育、防救災管理、公共建設等面向，進行土地使用分區、疏散路線、避難場所等規劃工作，亦可應用於都市擴張分析、傳染性疾病追蹤分析、治安犯罪斑點分析、交通易肇事地點分析、垃圾清運路線規劃、公共設施劃設、展店選址等，提供社會現狀分析、區位選址、設施配置等都市計畫參考之依據。

二、建議

本研究僅以視覺化呈現人口於各空間單元下分布的差異性，未量化數據探討

各空間單元的差異性與正確性，建議後續能以量化方式，探討各空間單元的差異程度，進而了解若未使用一級發布區與最小統計區等較細緻的資料進行規劃或研究時，將導致多少程度的誤判。

參考文獻

1. 黃任薇，2007，GIS 網格解析度之研究（未出版之碩士論文），國立成功大學都市計劃研究所，臺南市。
2. 張國楨、張文崧、曾露儀，2012，都市土地利用與人口老化對基層醫療資源分布之影響：以台北市為例，地理研究，第 56 期，25-40。
3. 張國楨、陳明勳、林嘉宏、王翊芬（2014 年 6 月），統計單元調整下空間差異性之研究。「2014 台灣地理資訊學會年會暨學術研討會」發表之論文，高雄君悅酒店。
4. 歐采瑀、洪榮宏（2014 年 6 月），領域統計資訊與統計區資料結合之影響因素分析。「2014 台灣地理資訊學會年會暨學術研討會」發表之論文，高雄君悅酒店。
5. Simon J. Sheather（2004）. Density Estimation. *Statistical Science*, 19(4), 588-597. ❖