

統計概念的啓蒙－統計圖

蘇國樑
國立空中大學 商學系

摘 要

本文重點在一般主體對資料集的統計認識過程，特別是在幾何化主體的統計認識概念，從而得到統計圖和統計概念之間的關係。其階段性的過程依序為(1)整體資料集的摹寫；(2)摹寫的簡化和符號的使用；(3)符號並列；(4)連續性條形圖的使用；(5)統計長條圖對資料集的統計概念的概括。另外，亦得出隨機變數在實踐上的意義是，不僅是對原始資料集內組成元素的測量活動和其結果，且其作用結果可以反映出原始資料集所擁有的現象和性質及統計結構。

關鍵詞：統計認識、資料結構、統計概念、統計思維。

前言

在探討統計教育或教學時不能忽略統計概念的發生和發展過程，因為教材的編撰和課程的安排應儘量符合主體統計概念的發展過程，這樣才不致產生教學與學習上的排擠效果。我們在編撰統計教材和課程時，可以依一般傳統教材的編排方式再加上舉例和解說。這樣的做法並無不可，但是傳統教材和課程是否根據了主體的統計概念發展而編排。如果是，則一般主體的統計概念和其發展的過程是什麼？如果傳統的編排方式是不同於一般主體的統計概念的發展過程，則更應該進行對一般主體的實際統計概念和其發展過程的研究。

統計概念的發展過程只有對一般主體或學生的統計認識進行實際的觀察才能獲得，而不是只以統計專家們的統計知識為主，且配合統計專家的觀點而構成的統計概念。因為統計概念並沒有區分一般人的、學生的或統計專家等的統計概念，如同十進位的加法就是十進位的加法，不能區分為小學生的、一般人的或數學家的十進位的加法。當然統計專家的統計知識亦是統計概念發展的重要研究對象，因為它們是規範和制約統計方法和理論的價值。也就是說，統計專家的統計知識可以用來判斷哪些統計方法和理論是值得發展和足以影響統計學科的方法和理論。然而，本文重點在於，統計教育或教學所關心的是一般主體對統計認識的過程，即一般認識主體的統計概念。因此，將分幾個方面加以討論：1.統計

概念幾何化的必然性；2.統計圖的意義；3.數量化資料集的統計圖化；4.隨機變數的統計意義；5.對統計教學的意義。另外，本文中的統計圖是以長條圖為說明對象，至於圓形圖或折線圖等其他的統計圖概念發展可列為將來研究的課題。

1. 統計概念中幾何化的必然性

近年來許多學者對統計思維(statistical thinking)¹頗有興趣，然而，統計思維是什麼？統計思維的對象是什麼？筆者認為統計思維的對象當然是資料集(a data set)，它們可以是一堆具體物資料集、一組數據資料集、一個有趣的現象或甚至是一個次數分配表等。但是，如果統計思維是以統計概念和統計概念彼此之間的關係為對象的思維運作過程，則統計概念在統計思維中是佔有重要的地位。資料分配表或統計分配表是以資料集內在的數值關係為內容所表現出的數量化統計概念，但數量化概念如何再形象化或具體化而能為感官可以直接感覺到的具體化概念？統計圖就是企圖利用幾何圖式或形式而給予主體對資料集的一個概括性的統計概念，但這一概括性的統計概念並不表示其為一簡單性的結果。統計圖亦是經過一番精確性和客觀性的轉化而成的幾何形式的統計概念，也就是說，它不是依一己之見而任意做出的結果，至少它必須具有能夠表現統計概念的作用和功能。既然統計圖在統計概念的發展和培養上具有概括的作用，則統計圖概念的發生和發展在統計教育上就是具有聯絡統計知識和統計概念的關鍵地位。總之，它是一種以幾何圖形來描述資料集的整體關係²，使資料集的內在結構視覺化，這亦是一種從抽象上升到具體的思維作用的結果。

2. 統計圖的意義

在《統計概念的啟蒙和發展》³一文中，對具體資料集的活動所建構而獲得的統計概念的第(4)階段中，是以資料集內的每一個子類為一個新的群或是新的測量對象，提到在各子類內進行重造表徵物活動或是數概念活動。例如，一組積木的呈現、說明或重製可以是以符號表之，譬如，X色是x個相同大小的“○”、Y色是y個相同大小的“○”、Z色是z個“○”、W色是w個“○”等；或採依序抓取或說明有X色有x個、Y色有y個、Z色有z個、W色有w個等。此時積木的共同符號表徵物是“○”，可以被替換成“√”或是“□”。³這裡值得探討的是何以符號可以用來表示具體物，並使主體獲得資料集的統計概念？站在知識論的立場而言，這也說明了統計概念的獲得是不需要藉助於數概念或數值關係。

對具體資料集的描述或解釋活動的最直接和有效的方式是，把具體資料集直接地、逐一地呈現在認識主體的面前。但是這樣的作法容易受到時空的限制，因為需要進行具體物

的遷移活動。這樣的作法也是無法產生統計概念的，因為連記憶或感覺經驗的使用都未曾出現。這是完全受制於客觀現實，同時也無法突破客觀現實的物質約束。

本文仍然是以「一次呈現一組相同外表形狀的、不同顏色的積木，做為實驗研究的具體物資料集。」並將探討從資料集內元素的代替符號的出現起，至長條圖的產生的一系列過程，研究統計圖和統計概念之間的關係。並參考拙著《國小兒童統計概念分析之研究(II)》⁴和其中的實例，歸納出 5 個階段性過程，依序為(1)整體資料集的摹寫；(2)摹寫的簡化和符號的使用；(3)符號並列；(4)連續性條形圖的使用；(5)統計長條圖對資料集的統計概念的概括。下面將依次詳述之。

第(1)階段，整體資料集的摹寫。這是利用對資料集的成員進行摹寫、畫圖的方式，也是對資料集的具象感覺的反映並加以“外化”，也就是說，主體掌握了資料集元素的外在特徵、形狀和其空間位置，並加以描繪成圖，譬如，畫出積木的輪廓、顏色和其所在的位置。這是一種單純的視覺感覺的反映，並轉化成直觀的或素樸的具體形象，然後將此素樸的具體形象畫出。這種具體形象的外化雖然是一種抽象作用的結果，但離不開對客觀具體物的直觀反映。因為每一種具體圖像都不是具體物本身，而是對具體物的抽象圖像。

對資料集的摹寫雖能十分具體地掌握資料集，但基本上這只能反映資料集的存在，卻無法產生統計思維的對象——統計概念。只是單純的直接反映作用，並沒有產生調節或控制資料集的作用。然而，統計概念是一種對資料集內在的結構關係的認識過程³，這是主體調節或控制資料集的活動，所引發的主體思維活動和作用的結果。因此，將具體的形象再簡化成符號不僅保留存在性，亦可做為思維的和思維作用的對象。思維對象是不受時空限制的。

第(2)階段，摹寫的簡化和符號的使用。以摹寫的具像反映為基礎尋找出更簡化的符號和敘述成員主要特徵來表示資料集，譬如，一個正方形“□”、圓形“○”或“√”來代表一個積木，並以文字描述顏色來代表實際塗色。這裡的困難在：一是，如果元素的大小形狀是不同時，如何以相同大小的符號表示不同大小的物件？然而，如果主體關心的是某類元素在整體資料集中存在的範圍或程度，則大小形狀和所佔的空間位置並不影響到元素的存在範圍或程度，這是對具體物的保留概念。而利用相同的符號配合文字對顏色的說明，再利用 1-1 對應的方式表示不同顏色的元素，則可獲得到“資料集”和“符號集”兩個同構關係的集合。這是將原來的具體物資料集轉化成“質和符號”所組合成的序對關係，這也是以具體資料集為基礎的客觀性符號集。另一困難是，每個主體所使用的符號並非一致的，因為每個主體有可能使用其自創的符號。然而，一個符號只能代表一個元素卻是一致的，因為受

到客觀性和存在性的制約。



資料集重現雖然不再停留在需要把原始資料集的成員摹畫出來，但仍然處於利用表徵代替圖形做出 1-1 對應的複製。這是把具體物資料反映成具體形象，再轉化成符號形象的體現結果。基本上注重的仍是對資料集的存在問題，因為符號是另一種資料集存在的表達形式。

第(3)階段，符號並列。如果主體關心的是以符號表現資料集的存在形式，則用以表達具體資料集的符號的排列形式是不會影響到存在的內容。也就是說，符號的重新安排只影響到符號的空間相對位置，並不影響所代表的資料集的組成內容和元素。這同時也表現出以符號所代表的資料集的內容的唯一性。然而，若各子類內符號元素做任意的堆積，則是無法觀察和比較出各類元素之間的多寡程度。因此，應將代表各類的符號在一個給定的基礎線上進行直線的、連續的並列，這樣才能比較或依視覺感覺出各子類元素按所占的範圍或程度。這種將表徵物做緊鄰的直線排列而進行累積，可依累積的長短程度來表現出該子類物件的多寡。例如，將用來表示 2 個子類內的表徵符號並列，令其分別為“✓✓✓✓✓✓✓✓”和“✓✓✓✓”，或是“□□□□□□□□”和“□□□□”，則很容易區分出哪個子類內部事物較多。這裡並無需運用到造數活動。這也是以離散表徵物或符號來表現離散的事物。

這種符號並列的形式來表現一組資料集，不僅可以表達組成資料集的各子類的內容的範圍或程度，且可表達資料集內容的存在性和唯一性。這種調節和控制符號的方式就是一種思維進化的結果。

第(4)階段，連續性條圖的使用。既然由觀察各子類元素的多少可以獲得資料集的統計的概念，則可以改用等寬的連續長條圖來代替直線併列的離散符號。這時長條圖的長度表示該子類內元素的多寡程度，如同用直線並列的離散符號的長度可以表示該子類內元素的多寡。如果我們希望獲得的是一組資料集的整體統計概念，而不是只詳細地列出各子類元素的個數，則可藉由簡單的條形圖符號所構成各子類元素多寡的統計圖，如長條圖，而求得到一個資料集的整體概括的統計概念，其中類組內元素的多少可以由圖形的長短得到相應的概括概念。這樣的幾何概括概念是一種了解資料集的概括的內在結構關係。

在以長條圖表現子類內部事物的多寡的關係時，需先選定共同的標準單位長度，如此才可利用長條形的符號表徵物之長短來表現子類內部事物的多寡。也就是說，在選定共同的單位長度量之下，可藉由連續的幾何圖形(如長條圖或直方圖)的長短表現子類內部事物的多寡的關係，取代了以共同的單位表徵物“□”的累積關係。更可以將其中不連續、有空隙的圖形更加美化和簡化。例如，

 和 。而此一連續的幾何圖形的長短必是所設定之單位長度量的整數倍的累積。因此，在運用連續的幾何量表現離散量時，雖然在視覺上所感覺的是一個連續的量，但實際上所表現的是一組離散量。

第(5)階段，統計長條圖對資料集的統計概念的概括。各子類內元素可用一種連續量的連續條圖來表現後，整個資料集才有可能由幾個連續性的條圖的組合所呈現出。這種“幾個連續性的條圖的組合”呈現的目的在對整個資料集內在結構和關係的整體呈現，即透過圖形表現整體的內在關係而使主體能做出整體的認識²。統計關係的統計圖形化是一種將事物內容和關係再具象化的表現，這樣的具象化表現並不是單純地將具體事物重現或以特殊的表徵物重現具體事物為滿足，而是以更一般化的幾何圖形來進行內部相互作用和關係的重現。也就是說，以一組統計的長條圖的形象來抽象和概括一組資料集的內部結構，其中的過程不僅由具體上升到抽象層次，更是一種再從抽象上升到具體的理性活動。

綜合上述的分析，統計圖可以呈現出資料集的統計概念的基礎，在統計圖可以呈現出資料集的內在關係或結構，而統計圖化的統計概念是主體透過分類活動，建立出的一系列的圖形化的序對關係。這些序對關係是由“子類和相應具體物的摹寫圖形”的序對關係，進化到“子類和相應符號的並列”的序對關係上，再進化到“子類和相應條形圖的長度”的序對關係。這就是一種由具體上升到抽象，再從抽象上升到具體的、具象的理性思維活動。

3. 數量化資料集的統計圖化

前一節所述的是關於具體物資料集結構的統計圖化，即利用符號或圖形來呈現和說明具體物資料集的內在組成關係。當具體物資料集內的元素特徵並不是可以藉著感官的感覺做出明顯的區分時，則需要藉助於具有共同的或熟悉的測量標準單位的中介物測量儀器，才能進行對統計對象和中介物的比較活動而獲得區分，也就是主體在對象和中介物相互作用的結果。例如，身高、體重或氣溫分別需要藉助直尺、磅秤或溫度計，才能測出對象物的身長、體重或溫度；又，例如，一個學生在一篇文章內的錯別字數，這是需要藉助於實際點計錯別字的字數，才能得知。

對目標資料集的元素進行測量活動的結果是一組分散的、個別的數量化資料集或數據集(a numerical data set)，這些是屬於數量化資料的蒐集活動和結果。測量活動的合適性當然會影響所得的數據，但這裡比較關心的是數據集的分類活動和其作用。雖然測量活動所得的量是一種質的表現方式，但它們除了分散的個別性外，不像顏色或形狀可以具體地被視覺感官所區分。

因為數據資料是主體對統計對象的某些質的特徵實施測量活動的結果，故數據值的來

源有連續性和離散性之分。所謂離散性數據資料集是指資料集內的元素的某項特徵(nature)是透過點計活動而得者。具有相同點計結果的元素即是具有相同的特徵，譬如，在一篇文章中發生相同的“錯別字數”的學生是視為具有相同“錯誤數”的特徵。這是將質的特徵轉化成數字形式，其中無法避免地會喪失一些重要資訊，譬如，“相同的錯別字數”並無法區分是否發生相同的錯字。然而，將質的特徵轉化成數字形式才能做進一步的分析。離散資料集的分類基礎相對而言較容易掌握，從而較為具體，因為代表元素特徵的數字不同即表示具有相異特徵的原始元素。因此，元素間的特徵較容易區分，從而容易進行分類和歸類活動。

所謂連續性數據資料集是指資料集內的元素的某項特徵是可透過連續性測量工具所測得者。相對地，連續資料集的分類比較抽象，因為元素所表現的特徵數字可能是連續性測量工具上某一個數字。數值間的差異是因為測量工具的使用結果，但原始元素間的實質上的差異有時感官是不易感覺到的，例如，30度和31度的氣溫差異是不易被察覺到的。因而數據值相差較小的元素間表其相應的實質特徵差異也小，數據值相差較大的元素間表其相應的實質特徵差異也大。

因為離散數據資料集的分類活動和具體物資料集的分類活動有許多類似之處，因此，接著將討論連續性資料集的分類活動。既然連續性資料的獲得是透過連續性測量工具所得，故個別元素的測量值相近者表其相應的性質或特徵相近；而個別元素的測量值差異較大者，則表其相應的性質或特徵差異較大。因而考慮將測量相近者歸入同一類，故子類內的數值差異不大，而相異子類間的數值差異較大。然而，這種分類方式對主體而言是非自發性的、不自然的，故不易掌握和自創。因此，通常需要透過學習獲得以全距為基礎，將全距分成幾個子區間而獲得分類的標準，再進行分類。這裡所謂的全距是資料集中最大的數值和最小的數值所涵蓋的範圍，此一範圍也包含了資料集內所有的數值。所以，對數據資料集做分類不僅是對數據集的形式做分類，且是對數據集的實際內容的質特徵做分類。也就是說，不論是具體資料集或是數量化資料集，分類的基礎都是建立在資料集的元素特徵上。

無論原始資料的連續性與否，測量活動所得的結果數據資料集皆是一些分散數值，正因為此一分散性或個別性主體才能進行分類活動。而分類活動後才能進行各子類內元素的並列活動，甚至可以在元素和符號之間建立對應關係。繼而以統計圖呈現出對應的資料集結構，即資料集的統計概念。分類活動是以量的差異為根據來區分質的差異。雖然分類的標準在連續性資料上只能做原則性的規範，然而，分類的基礎也需要符合原始資料集的特

性和客觀性。例如，醫學上使用藥劑量差異的界定就屬於比較敏感而細微；人體對氣溫的感覺差異的界定就比較不敏感而寬廣。

4. 隨機變數的統計意義

前節談及數據資料集是主體對統計對象的某項特徵所做的測量(measure) 活動或點計(counting)活動所得的結果，這在數學上所隱含的意義是，將資料集(Ω)內所有的元素對應至實數集(real number set, \mathfrak{R})的一個對應關係，此一對應關係通常記作“ $X: \Omega \rightarrow \mathfrak{R}$ ”，其中的“ X ”在統計上稱作隨機變數。因此，隨機變數就不僅隱含著將“元素特徵”的質轉化成量的作用和關係，且隱含著作用在對象特徵上的測量實踐活動。如果離開此一實踐活動隨機變數將成為毫無統計意義的純粹對應關係，也使得隨機變數僅留形式而失去實際可測活動的內容，從而使隨機變數停留在抽象層次。

由於隨機變數包含這一層實踐測量活動，因此，可以進一步的深化它的作用和功能，即不僅是一個將原始資料集元素對應到實數的作用，還包含了造就出另一組可用“量的關係”來表現原始資料集內元素的“質的差異關係”，和其所屬的資料集內在的統計結構。也就是說，隨機變數呈現了一個序對關係，即“元素特徵的質的差異”和“對應於資料集的統計結構”。換成統計術語就是隨機變數和其相應之次數分配模型所成的序對關係。這裡的次數分配模型包含了圖形化的模型。這也是隨機變數和一般對應關係不同之處，因為一般對應關係是用來維持被對應物和對應物之間的對應關係，譬如，一般的函數對應關係只是從一個數集到一個數集的對應關係。然而，隨機變數卻不僅是用來維持被對應物和對應物之間的對應關係，且還需在對應關係下維持原始資料集的統計結構。

既然隨機變數是可以反映對原始資料集的測量活動和其結果，則隨機變數自然受到原始資料集或原始統計對象的性質的制約，故其作用結果應能反映出原始資料集或統計對象所擁有的現象和性質及統計結構。這也使隨機變數從抽象的對應關係再上升至具體的層次，形成抽象和具體的統一。因為統計結構不應在實踐的測量活動的過程或結果中喪失或改變，否則此一實踐活動隨即失去其客觀性和實用性。

5. 對統計教學的意義和結論

本文不但把有關統計圖概念的發展過程具體化，且把其階段性模式化。這也是主體藉著整體認識資料集來對客觀事實進行了解的具體化過程，由此過程中可以看出人的認識是創造性和多樣性的。這樣的結果一來，較有可能做為統計教學中教材的編撰和教學方式的根據，二來，避免陷入過去可能的僵局，就是缺乏學生的統計概念的發展階段和程序，而

無法給出適當的協助。這是本文研究和開發統計概念的實用性的具體表現。

Garfield 曾提出「統計概念的探討是未來統計教育學者所應思考的重點」⁵，且統計教育能否成爲一個獨立學門的重要基礎在於統計概念的特殊性和其特殊的規律的研究。由於一般學者常以數學概念爲基礎來研究、探討統計概念，故會產生如 Garfield 所提的「常發生對學生學習統計的研究結果和心理研究互相矛盾」⁵。由統計圖概念的發生和發展的過程亦可看出，數學概念和統計概念是有其相異之處，然而，這個相異之處正是促成彼此間的相輔相成的基礎，例如，數理統計的發展可以豐富數學理論的內容。這是本文研究和開發統計概念的實用性的另一個具體表現。

在第 3 節探討、分析量化資料集的統計圖化的過程中，引發筆者對隨機變數的重新思考。過去對隨機變數的解釋比較偏重“對應關係”，如 McClave 和 Dietrich 所定義：「隨機變數是對一個實驗中的一個簡單事件給與一個數值的規則」⁶，因而容易忽略其背後隱含的測量的實踐活動，這個重新思考不但恢復隨機變數應有面貌和作用，亦是本文實用性的第三個具體表現。

未來將繼續思考，統計次數分配的結構如何擴大、提升爲理想化的機率分配結構，則隨機變數不僅由原始的可測關係提升爲機率分配關係，這亦將可讓機率理論成爲具有實踐對象的理想統計結構關係。這也是統計科學中可以直接利用數學運算和邏輯推理，求出有關隨機變數的一些理論和性質的重要原因之一；而無需屢屢回顧直觀經驗的原始資料集的現象和性質。例如，經驗提供我們：學生 IQ 成績呈常態分配，因此，研究學生的 IQ 成績時，我們可以利用常態機率分配的各種性質來描述、推論或整體認識學生的 IQ 成績，而不是只能針對某一特定的學生 IQ 成績集合做出描述和說明。

註釋

1. C. Mallows：〈The zeroth problem〉，《The American Statisticians》，1998年，Vol. 52，No. 1，第1-9頁。
2. 蘇國樑：〈做爲整體認識的統計學〉，《數學教育學報》(中國)，8卷，1期，1999年。
3. 蘇國樑：〈統計概念的啟蒙和發展〉，手稿，1999年。
4. 蘇國樑：《國小兒童統計概念分析之研究(II)》(NSC84-011-S-180-002N)，1995年。
5. J. Garfield：〈How students learn statistics〉，《International Statistical Review》，1995年，Vol. 63，No. 1，第25-34頁。
6. J.T. McClave 和 P.G. Benson：《A First Course In Business Statistics》，Dellen Publishing Company，a division of Macmillan Inc.，San Francisco，1989年。