

統計對象的抽象原則

蘇國樑
國立空中大學 商學系

摘要

本文藉著探討統計方法和統計思維的過程，歸納出 3 個統計對象的抽象原則：一、資料結構的序對化原則；二、資料結構的單值化原則；三、資料結構的函數模型化原則。這 3 個抽象原則不僅符合統計思維的發展過程，亦符合了人類認識的思維規律。因此，在教學上這 3 個抽象原則不僅是建立統計方法的基礎，而且可以在統計方法和統計思維之間起著聯絡和指導作用。

關鍵詞：資料結構、統計對象、統計抽象、統計認識、統計等價。

C. R. Rao(1998)曾說：「位置和分佈散差的測量，高階動差和指數的度量，並通過某些圖形，如直方圖（次數分佈），條形圖及二維平面圖來表現數據的醒目特徵。這些方法並不涉及觀測數據的隨機結構（或稱為機率分佈）。因此，計算得到的各種描述統計量可用來比較不同的數據集合。」這樣的說法其實就已經涉及觀察數據集或資料集的隨機結構，因為數據的紀錄就是為了了解隨機現象所做的觀察、測量和記錄的結果。所以，次數分佈或機率分佈都是為了探究帶有隨機現象的數據集的結構。然而，如何才能探究或認識帶有隨機現象的數據集的結構，這是統計學的重要課題。本文首先將探討統計方法的重要性。其次，說明統計抽象的意義。接著以 3 個方面來討論統計對象的抽象原則：一、資料結構的序對化原則；二、資料結構的單值化原則；三、資料結構的函數模型化原則。接著是說明這 3 個原則在統計教學上的意義。最後發現這 3 個原則不僅符合統計思維的發展過程，亦符合了人類認識的思維規律。

這裡須要先說明的是，統計處理或分析的對象是資料集，它可以數量（據）化的也可以是非數量化的，故本文均稱之為資料集(data sets)。因為資料集的數量化是需要透過隨機變數的轉化，不過這種轉化的本質「仍是為了說明資料集內元素彼此之間的質的差異」（蘇國樑，1999c）。因此，數量化的或非數量化的資料集都是一種記錄元素或個體的質的結果。

一、統計方法的重要性

C. R. Rao(1998)提出：「基於數據集的特徵和要解答的問題，甚至於制定了一些規則來

選擇一些可替換的統計量，如描述分佈位置的平均值、中位數和眾數。這樣的統計分析被稱為描述數據分析(descriptive data analysis)。另一方面，在理論統計學中，雖然其目的也是歸納數據，但是主要是強調作為機率分佈基礎的一個特定的分支（或稱為模型族）。在這種情形下，歸納或記述統計量主要依賴於某個特殊的機率模型。在推斷某些未知參數時它們的分佈被用於決定其不定性的範圍。這樣的方法於是被稱為是推論數據分析(inferential data analysis)。」這段話也就說明了不論是用分佈位置數（集中趨勢量數）、離中趨勢數或是機率（分配）模型都是為了虛擬、描述或推論數據集。然而，虛擬、描述或推論資料集的基礎是建立在資料集的內在結構關係上，因為主體是透過這些結構關係來重造、複製或虛擬出該資料集。因此，掌握資料集的內在結構關係就是掌握了統計概念。也就是說，找出資料集的分佈位置數、離中趨勢數或是機率（分配）模型等，都是為了建立資料集內在結構關係的方法或原則，即統計概念，而分佈位置數、離中趨勢數或是機率（分配）模型等都是統計概念的表現方式。本文所提到的主體是包括了學生、統計研究工作者或統計教師等。

既然統計概念是關於資料集內部結構關係的概念，從而統計思維就是以統計概念為基礎和對象的思維運作活動（蘇國樑，1999c）。統計思維的規律和人類思維的規律應有某些一致性和關聯性。C. R. Rao(1998)在《統計與真理》一書就曾提到：演繹法、歸納法和誘導法的思維方法。然而，這些是一般推理和思維的方式。對統計而言應有其特有的思維方式或概念處理的方式。例如，談到科學方法就有所謂的歸納法、演繹法和類比法等。又例如，數學求解問題時的思維方法中就有所謂的化歸原則和關係映射反演方法；數學抽象的方法論原則就有弱抽象的特性分離一般化原則和強抽象的關係定性特徵化原則等（鄭毓信，1991）。然而，如何獲得這些特有的統計思維方法和原則呢？因為統計學家是利用統計符號或是符號運算式，表達並記錄下他們對資料集所做出的統計思維和認識。因此，可以藉由探討統計學的歷史發展過程或是探討統計學家為了解釋、描述統計對象所做出的統計活動和其紀錄，從而找出統計的歷史發展規律或是統計的活動過程規律。

二、統計抽象的意義

統計學是透過分析資料而達到解釋資料的科學，所以，統計又是一門資料分析的科學(a discipline of data analysis)。C. R. Rao(1998)說「統計學沒有任何固有的對象，是一門獨特的學問。依賴於由解決其它領域內的問題而存在並興旺發達的。」這句話的意思應該是指統計對象的“內容”不必是特定的、指定的，可以是非特殊的或任意的。這樣的統計方法和理論才具有普遍的適應性，也才能解決“其它領域內的問題”。但是，統計對象必須是固定的，

否則會造成缺乏統計認識的對象，從而無法進行統計認識或思維活動。這裡所指的統計認識的對象是包括了資料集、統計圖表、統計符號或統計方程式等客體。

C. R. Rao(1998)所提的“其它領域內的問題”是指有關其它領域內的“資料分析問題”。然而，資料分析的意義是什麼？最基本的統計分析方法是對資料集進行分類和點計活動，再透過分類和點計活動所得的結果再對原始資料集進行了解、解釋、重造或推論。就是說，透過統計分析的過程達到對資料集的統計認識。因此，分類和點計活動所得的結果是認識原始資料集的基礎，因為主體不容易直接地理解或認識無序的原始資料。然而，分類和點計活動所得的結果並非原始資料集，而是對原始資料集的重新整理或轉化。譬如，資料處理過程：原始資料→次數分配→平均數，其中次數分配和平均數都已經不是原始資料，有時候它們和原始資料也相去甚遠。換句話說，就是變得抽象了，這就是一種統計抽象化的過程和其結果。

主體透過對資料集的分類和點計等統計活動和其結果，可以獲得並建立「資料集的結構」或「資料集內在組成元素的關係」（蘇國樑，1999a），而這些抽象化的結構和關係不僅是認識統計原始的或具體的資料的基礎，也是做為進一步統計認識和分析的對象。因此，統計抽象化是統計發展的必然過程。進一步的抽象化的意思是，透過邏輯的或數學的分析和操作抽象後的結果所得出的一些性質。

三、統計對象的抽象原則

人類思維不易掌握雜亂的、無序的對象，故需藉由實踐活動中建立出一些認識客觀對象的規律和方法。這並不表示透過認識的規律和方法的活動所得的結果和原始的認識對象是一模一樣的。而是透過認識這個已經被整理過的對象去認識原始對象的概念，因為整理過的結果對象是較原始對象有“秩序”的和“規律”的。這就隱含著整理過的結果對象和原始對象之間存在著某種等價關係，譬如，資料處理過程：原始資料→次數分配→平均數，在統計上用次數分配來“表示”或“說明”原始資料，或是用平均數來“代表”或“描述”原始資料或次數分配。這種可以在兩個不同的統計對象之間進行“表示”、“說明”、“代表”或“描述”，就表示其中存在著某種等價關係，可以稱之為統計等價(statistical equivalence)。

統計概念上的“等價”和數學概念上的“等價”是有差異的，統計等價並不用來表示兩個統計對象之間是相等或是全等，而是透過“表示”、“說明”、“代表”或“描述”來表現兩個統計對象之間的某種相似性、共同性或替代性。這也是統計和數學的差異之一。主體可以透過認識這種等價關係從而認識原始的統計對象。接下來將探討透過這種統計等價關係從而獲得統計抽象的原則。

(一) 資料結構的序對化原則

統計圖表（譬如次數分配表或統計圖）是一個可以提供主體認識資料集或群體的活動過程和其結果。雖然統計圖表不一定可以找出一個唯一與之對應的、客觀存在的具體群體，但在現實界中面對一個客觀資料集或群體時，主體卻可以找到、做出與之對應的統計圖表。主體是透過製作統計圖表的活動過程中，獲得和認識資料集或群體的內在關係，並將此一內在關係具體化成統計圖表。因此，透過統計圖表可以認識相應的資料集或是群體。故在形成統計圖表的思維和活動的過程中，提供了主體認識資料集或群體的一般性和原則性的方法和其結果。

既然統計圖表是一種認識資料集或群體的結構的一般性方法和其結果，則建立統計圖表的基礎便是一種具有方法論的思維原則。事實上，能夠建立統計圖表的基礎在於掌握住資料集內在的序對關係，這是對紀錄性資料集的一個更進步的思考。這就是所謂的資料結構的序對化原則。它是透過分類和點計活動來獲得和建立資料集內在的類、屬性或特徵和其所屬元素所涵蓋的範圍或程度所成的序對關係，這個序對關係就是資料集內在組成元素的“質和量”的序對關係（蘇國樑，1999b）。譬如，若資料集為 $\Omega = \{\circ\triangle\circ\square\triangle\circ\square\triangle\square\square\circ\circ\triangle\circ\}$ ，經過分類所得的序對關係之一是 $R = \{(\circ, 6), (\triangle, 3), (\square, 4), (\diamond, 1)\}$ 。通常我們利用 R 來表示和描述 Ω ，這就表示 R 和 Ω 之間存有一種統計等價的關係。主體便可以透過此一序對關係來獲得和建立對原來資料集的有序的概念，並做為建立統計圖表的基礎。因此，統計圖表就是將這種“質和量”序對關係的形式化和具體化，從而透過認識統計圖表可以獲得原始資料集的有序的整體概念。這透過統計活動建立內在組成關係或是資料結構的最基本和最關鍵的方法和原則之一。

統計圖表是直接對直觀認識對象所從事的認識活動而得的結果和其紀錄。這是一種直觀經驗的抽象所得的結果關係，故需受制於直觀經驗。雖然如此，統計圖表仍是一個最普遍的、直接的、有效的和準確的認識資料集或群體現象的內在組成的結構。這就是說統計圖表提供了主體認識資料集或群體的統計概念。如果資料集是全部的群體現象的紀錄，則統計圖表提供主體對客觀群體的整體概念。如果資料集是群體的部分紀錄資料，則統計圖表提供主體虛擬客觀群體的一個縮影。因此，倫敦商人 John Graunt 才能成為「用實例展示如何利用統計學來描述問題的現狀，並引導事物未來發展方向的人。」（C. R. Rao, 1998）

(二) 資料結構的單值化原則

要說明或比較兩個數集合之間的大小關係是一個比較麻煩的，但要說明兩個集合的包含(\subseteq)、交集或聯集關係則比較容易。例如，當我們定義“集合 X ” \leq “集合 Y ”的意思是“ $x \leq y$ ，

$\forall x \in X, \forall y \in Y$ ”，簡化為 $X \leq Y$ ；在此一定義下，若有“ $x > y, \exists x \in X$ 或 $\exists y \in Y$ ”，則無法結論“集合 X ” \leq “集合 Y ”。因此，若 $A = \{x \mid 0 \leq x \leq 2\}$ ， $B = \{y \mid 3 \leq y \leq 6\}$ ，則可以有 $A \leq B$ ；但是若 $C = \{x \mid 0 \leq x \leq 3\}$ ， $D = \{y \mid 2 \leq y \leq 6\}$ ，則雖可結論出 $C \cap D = \{x \mid 2 \leq x \leq 3\} \neq \phi$ ，卻無法結論出 $C \leq D$ 。具體地說，數集合之間的大小關係是不易藉由單純個別元素的比較而界定的，而是需考慮由一群元素和一群元素之間的整體比較。

除了數據資料集(numerical data sets)之間的大小關係是不易藉由比較資料集內在元素的大小而界定的；也無法藉由直接比較統計圖表而獲得資料集之間的大小關係。除了無法確定對資料集之間的大小關係外，統計圖表亦無法對資料集做出一個概括性的、總結性的思維。因而，象徵或代表一組資料集的一個事物、符號或數字將取代可以表現資料集整體結構的序對關係。這也是把資料集整體結構的序對關係的概括思維活動的再抽象化。

因此，統計的思維活動的另一特徵在建立資料集的單值化的思維原則。就是企圖利用幾個簡單的數值來代表整個資料集的結構，具體的說是希望利用幾個簡單的數值進行描述、解釋、概括或預測整個資料集的結構，或是更大範圍的整體現象的結構。這些數值稱為資料集的特徵數或參數。例如，以平均數或中位數來描述一組資料集，或是以平均數和標準差來虛擬一組資料集，甚至以一個相關係數來概括兩個具有統計關係的特徵、屬性或變數之間的「直線關聯的方向和強度」(David S. Moore, 1998)。這就是一種“資料結構的單值化”，也是以一個或幾個數值來總結一組資料集的結構和關係。

因此，資料結構的單值化原則的功能就是，透過數據資料集單值化所得之量數做為資料集的代表數，從而可以將量數的大小關係“視為”資料集之間的大小關係，或說明資料集涵蓋之範圍或程度等關係。例如，透過比較兩個平均數的大小關係來“表示”兩組資料集的大小關係；而透過比較兩個變異數的大小關係來“表示”兩組資料集之涵蓋範圍或離散程度的關係；又，在變異數分析中，透過比較幾個不同母體的平均數的差異來“代表”不同母體之間的差異。這樣的作法就是統計等價在資料分析中所發揮的功能和作用。

雖然資料結構的單值化有其簡單、方便之處，卻不易顯明地表現出資料集的整體結構，而且也不容易找出相應的具體對象來進行匹配。例如，欲虛擬出一個平均數為 3 的母體，可以有 $N(3, 1)$ 、 $N(3, 4)$ 或 $N(3, 9)$ 等無限多的可能。然而，正因為資料結構單值化的簡單和籠統，從而可以獲得一個概括性的整體概念。進而提供了主體認識資料集或群體結構的一般性和概括性的方法和結果，故建立參數或統計量數的基礎便是一種具有方法論的思維原則。

(三) 資料結構的函數模型化原則

在處理量化的資料集時，雖然單值化的結果可以簡單化一組資料集的整體成爲一個數字概念，但是卻無法讓主體認識一組資料集的整體結構，因爲它過於籠統和抽象。因而，將一組資料集的整體性結構或序對關係轉化成一種函數關係，從而建立其相應的統計結構模型。這是一種從抽象思維再進化到具體思維的過程。

雖然統計圖表也可以表現出一組資料集內在組成結構的整體模型，但這是一種直接建立在具體材料或直觀經驗現象所得的反映模型。使得直觀經驗的統計圖表所呈現的整體模型的客觀性受到時空的限制，從而無法符合科學規律中的一般性、普遍性和精緻性。因此，我們需要從直觀的經驗材料並配合客觀現象中的規律，歸納出或抽象出一個具有一般性、普遍性和精緻化的原則的理想整體結構模型，稱之爲理論的次數分配模型或是機率分配模型。

總之，資料結構的函數模型化原則就是：企圖利用函數關係（包括圖表關係）或數學方程式來代表整個資料集內在的“質和量”的序對關係或是結構。具體的說，是希望透過函數關係或方程式進行描述、解釋、或概括整體資料集的結構，從而推論出更大範圍的整體現象和結構。或者說，這些函數關係或方程式是統計等價於整體資料集的結構。例如，以常態分配的模型來表現對稱的鐘型分配的資料集結構，或以二項分配的模型來概括二值結果的資料集結構，甚至以核密度函數(kernel density function)(Su & Taylor, 1991)代表任意的資料集的整體結構。這些都是一種“資料結構的整體化”，也是企圖以一個數學函數形式來總結一組資料集的結構。

數學和統計學在某種程度上都可以看作是建立模型（簡稱爲建模）的科學，然而，數學上的建模和統計學上的建模是有區別的。數學上建模的基礎對象是自然現象的規律的形式化；統計上的建模的基礎對象是關於自然現象、人文現象或其紀錄的整體現象的組成結構。因此，數學建模側重在機械論的決定模型(deterministic models)，因爲視整體爲單一現象或事件所構成，發生不一致是由於觀察或測量誤差所致；統計建模側重在偶然性的隨機模型(stochastic models)，因爲現實中的一個整體可以由多個不同的現象或事件所構成，包含了誤差現象。爲了認識整體現象，統計建模需要考慮構成整體的內在關係模型，即理論的次數分配模型。故所謂的隨機模型應是現象或資料集整體組成的整體結構的模型，或稱爲整體模型(holistic models)。

四、對統計教學的建議

初看之下，統計對象的抽象原則和一般處理資料的統計方法有些相近，但是，這裡要探討的是處理資料的統計方法的發生原因和過程。這就涉及統計認識和統計思維，也就是

需要探討統計抽象的原則。因為統計方法也好、資料處理方法也好都是統計思維的結果。譬如，建立資料的次數分配的基礎在找出資料的序對關係，從而構成資料的統計結構，進而構造資料的統計概念。而完成序對關係的統計方法就是分類活動。因此，分類活動並不是只有單純的分類和歸類的動作，而是具有更積極的統計意義。如果只是要找出整組資料集的組成關係，是不需要透過分類活動，因為透過對資料集內容成份的逐一觀察、記錄也可以得出整組資料的組成內容。然而，這是一種低效率的作法。因此，在統計教學中就可以強調分類活動和序對關係在統計上的意義和作用。

其次，計算平均數、變異數或是迴歸直線等並不需要利用高深的數學知識和技術。但是，如果只是強調單值化的數學操作技術，不僅會陷入數學簡單化的缺點，從而忽略了單值化的統計意義和作用。因此，在統計教學上，單值化原則可以強調其做為描述、解釋和虛擬所對應之統計對象（資料）或可預見的對象（母體）的統計意義。還有單值化所產生的特殊統計功能，譬如，比較兩個或是多個群體之間的差異是單值化結果的功能之一。

第三，結構函數模型化原則更是希望能將資料的整體結構或統計結構，進一步的藝術化、精緻化、一般化和普遍化。因此，透過對資料的來源特徵和次數分配所建構出的機率分配模型，雖然是以數學方程式的形式表現出，但這不僅是單純的數學函數關係，而且可以表現出統計對象（樣本或群體）的結構關係。這就是結構函數模型化原則在統計上的意義作用。因此，在統計教學上除了介紹各種機率方法模型，還要強調如何在資料的次數分配的基礎上建構出對應的可能的理論機率模型。

最後，對於熟悉資料處理的統計方法的人而言，這些統計方法是簡單的、具體的和有意義的，然而，這些統計方法的統計意義卻是建立在上述的三個抽象原則上。因此，教學上就必須隨時將統計方法和其對應的統計意義之間做出聯絡。資料處理的統計程序也都是符合上述的三個抽象原則，因為這三個抽象原則是認識統計對象或是獲得統計概念的基本原則。因此，在統計教育上統計方法可以看成是認識統計對象或是獲得統計概念的操作形式，而統計對象的抽象原則就可以看成是統計方法在處理資料的思維內容。

五、結論

不論是單值化或整體化不外乎希望簡單化、明確化和顯明化地將資料集的結構一次表現出來，使得主體能一次或迅速地掌握住資料集的組成結構的整體概念，即資料集的統計概念。其次，透過上面關於統計抽象原則的論述，可以看出統計抽象原則是為了獲得統計概念和其關係的思維活動，而統計方法就是這些思維活動的結果。因此，在統計教學上除了強調統計方法的傳承，還要強調統計思維的傳承。

在第三節中所談的不僅說明了統計認識的思維原則，也說明了機率分配模型建立的過程。雖然在理論上和實踐上，資料結構的函數模型化和資料結構的單值化兩者都有求參數的過程，但是兩者的基本立場是不太相同的。資料結構的單值化中求得特徵數的基礎背景在，企圖利用一個或幾個數值來表現一個資料集的全體，或是說希望藉由一、兩個數值虛擬出一個資料集的來源母體。這會造成不同主體之間有著不同的統計概念和判斷，因為不同的主體在針對同一個統計特徵數可能虛擬出不同的結果對象。譬如，欲虛擬出一個平均數為 3 的母體，可以有 $N(3, 1)$ 、 $N(3, 4)$ 或 $N(3, 9)$ 等無限多的可能，這是過於簡化的結果。然而，資料結構的函數模型化的基礎背景在，企圖利用一個函數模型來表現或實擬出一個資料集的來源母體。這是希望求得一個整體的認識，而不只是整體的虛擬，也是企圖將一個未知母體的整體結構再具體化。

在探討統計思維方法的過程中，可以看出統計思維的模式(pattern)是從序對化關係的建立過渡到單值化的整體概括概念，再由單值化的概括概念過渡到整體結構模型的建立。這正符合了人類認識思維的發展過程和規律：由具體思維上升到抽象思維，再由抽象思維上升到具體思維。因此，在統計教學中就可以透過學生對統計方法的操作，從而了解學生的統計思維活動，進而了解學生的對統計概念的掌握程度。因為統計認識和統計思維的規律符合上述的 3 個統計抽象原則。

六、參考資料

1. Rao, C. R.. *Statistics and Truth: Putting Chance to Work*. 石堅、李竹渝譯(1998)，《統計與真理》，台北：九章出版社。
2. 蘇國樑(1999a)：做為整體認識的統計學。《數學教育學報》，8(1)。
3. 蘇國樑(1999b)：統計概念的啓蒙和發展。《科學教育月刊》，220, 9-16。
4. 蘇國樑(1999c)：統計概念的啓蒙和發展(II)——統計圖。《科學教育月刊》，221, 2-9。
5. 鄭毓信(1991)：數學方法論。中國：廣西教育出版社。
6. Moore, David S.. *Statistics*. 鄭惟厚譯(1998)，《統計，讓數字說話》，台北：天下遠見出版有限公司。
7. Su, K-L (蘇國樑) & Taylor, R. L. (1991). *Consistency And Rates of Convergence for Failure Rate Function Estimators*, Technical Report No. 151, University of Georgia.