# 什麼是加權？

## 數據分析淺談

近日，有政黨就一政治議題進行電話普查，結果引來抨擊。有評論指該調查未有進行加權（weighting），故質疑其可信性。[1] 筆者希望透過本文，讓讀者明白加權對普查的意義，以及其限制。

### 加誰的權？

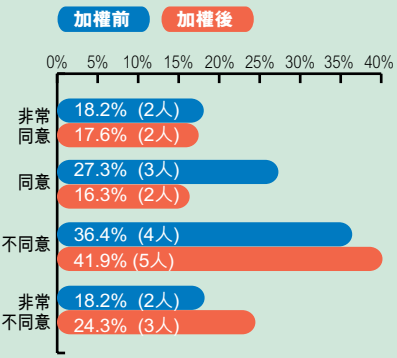
加權是指改變某些數據值的比重。當研究員發現太多或太少抽取某些羣體的數據，便須進行加權，使調查能夠符合現實狀況。普查的對象是社會大眾（如：全港市民）。理論上，除非由政府主導，否則一般機構難有足夠資源，為一份研究挨門逐戶進行訪問。所以基於現實考慮，一般研究機構都會以隨機抽樣方式，訪問特定數目的市民，然後以那些受訪者的意見推算出社會的普遍現象。但是，既然是抽樣，便會有誤差。誤差可以有許多類，其中一種就是關於抽樣受訪者的背景資料。

受訪者背景包括其性別、年齡、教育程度、出生地點、宗教信仰等。這些背景資料對研究結果可能舉足輕重。舉例說，假定男生愛看科幻小說，女生愛看言情小說。在全港中學生消閑讀物的調查中，若有六成受訪者是男生，我們可能得出「中學生愛閱讀科幻小說」這結果。然而，這結論很可能純粹因為受訪的男生比例過高，以致研究結果偏向男生的選擇。有見及此，我們需要透過加權減低性別對數據造成的傾斜。

### 誰加誰減？

加權一般做法是：將比例過多的羣體用一個小於1的值來加權，將比例過少的群體用一個大於1的值來加權，加權值是實況及研究中的比例。假設正常男女比例為2：3，但研究中男女比例為3：2，男性的加權值是2/3=0.67，女性的加權值是3.2=1.5。因此，研究中所有男性的答案會調節至原先的三分之二，而女性的答案將按比例上升一倍半（3/2=1.5）。表一是加權前後的數據分布：

表一　題目：我有責任維護弱勢社羣利益



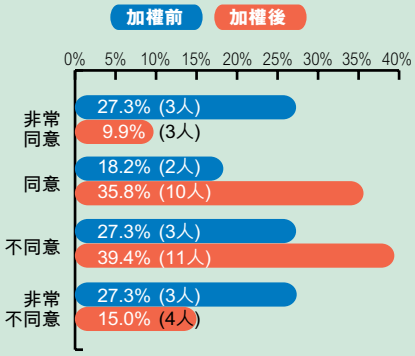
林俊杰
突破事工研究及發展部研究幹事

加權前，「非常同意」及「非常不同意」比率相等，且是選項中比例最低；加權後，「非常同意」及「非常不同意」的比率改變了，卻不再是選項中最低。由此可見，加權與否對研究有一定程度的影響。

## 不加權，行嗎？

若研究旨在描述社會普遍現象，加權是必要之舉。從表二的例子，你大概能明白加權前和加權後的分別。由於普查大多是抽樣訪問，若某群組的比例過大，便會使研究偏向某群組的意見。有人可能擔心，將受訪者數據加權會扭曲答案，破壞其真實性；但從另一角度看，其實一個500或1,000人的訪問有何代表性？惟有透過加權，讓抽樣結果更貼近現況，致使更準確掌握社會大眾的意見，顯出研究的價值。

表二　題目：你認為傳媒從商業角度考慮編輯（方向？）是合理的




## 加權不是天山雪蓮

然而，加權不是萬靈丹。正如本文開首提及的政黨民調，即使將其數據加權，那份問卷的代表性仍然存疑。原因是它的年齡羣組分布太偏頗。若調查裏羣組比例與現實情況差距太大，加權後的結果可以有翻天覆地的變化！以表二為例，加權前較多人選擇「非常同意」；加權後，這倒成為了比例最少的選項。

加權是統計學上合理的做法，不過當受訪者的背景資料與現實出現太大距離，便會削弱了加權的意義。面對這問題，調查機構便應該考慮檢視抽樣方法，加以改善，然後重新收集數據，方為合情合理之舉。

1 詳參區家麟：〈民建聯民意調查三處死穴〉，《主場新聞》，2013年4月9日及〈民建聯調查：七成人不支持佔中　受訪六成中老年　學者：沒代表性〉，《明報》，2013年4月10日。