淺談民調加權

文 | Roger Kuo, 益普索專案經理, Ipsos Project Manager

每當台灣選舉或重大政策民調公布時，新聞報導最後總會附上一行小字：「調查結果依據內政部最新人口資料，依性別、年齡、教育程度及居住地進行加權」。然而，許多民眾甚至政治評論員仍然對加權後的統計數據產生懷疑，甚至認為民調機構可以透過加權來調整數據，操控民調的結果。

事實上，加權（Weighting）並不是操控數據，而是統計學中用來修正樣本結構偏差的標準校正程序。在現代民意調查的實務運作中，透過調查前的隨機抽樣及調查後的統計加權，能讓調查機構在維持機率抽樣原則的前提下，有效控制執行成本。要理解台灣民調在實務上如何運用這項技術，必須從抽樣方法的選擇與台灣電訪環境的限制說起。

一、為什麼不直接在抽樣時「按比例抓人」（配額抽樣）就好？

很多人會問：「既然知道台灣 20 至 29 歲年輕人佔 15%，為什麼不在打電話時，規定年輕人滿 15% 就停，一定要先隨便抓再事後加權？」

在統計學上，這種「額滿即排除」的做法稱為「配額抽樣（quota sampling）」。台灣專業民調機構之所以捨棄配額抽樣，改採「無配額隨機抽樣及事後加權」，背後有兩個關鍵原因：

1. **配額抽樣會破壞「隨機性」，無法計算抽樣誤差**

   真正的機率抽樣要求「母體中每個人被選中的機率都是事先已知且可控的」。一旦設定配額，隨著「愛接電話的人」率先把名額填滿，後面才被抽到的同齡者入選機率瞬間歸零。這種人為截斷機制會引入嚴重的「可近性偏差」，使民調失去統計推論的正當性。

2. **在電腦電訪時代，設配額反而更貴、更浪費**

   傳統教科書認為配額抽樣比較省錢，但那是紙筆面訪時代的舊觀念。在現代電腦輔助電話訪問（CATI）中，系統自動撥號的邊際成本極低，「接通即有效」才是最省錢的做法。若設定年齡性別配額，專案後期為了抓到最後 5% 的冷門族群，訪員可能連打 30 通電話卻因「配額已滿」被迫掛斷，產生極高的篩選空轉成本。

因此，台灣民調機構普遍採用以下方式來進行全國性市話民調：利用市話區碼天然綁定行政區的特性，如此可以在「地區」上進行相對較低成本的「分層隨機抽樣」；而在「年齡、性別、教育程度」上則完全不設限，接通就訪，最後再以加權的方式來拉齊母體在「地區、年齡、性別、教育程度」的比例。

二、現實的困境：當隨機撥號遇上「年輕人消失」

然而，就算抽樣程序完全隨機，市話電訪依然面臨一個巨大的結構性難題：收回來的原始數據（raw data），長輩總是特別多，年輕人卻嚴重不足。

這背後的成因包含兩個層次：

1. **反應偏差（non-response bias）**即使家裡有裝市話，年輕人也經常不在家，或聽到市話響了懶得接，最後走過去接起電話的幾乎都是退休長輩。雖然部分機構會採用「戶中抽樣」（請接電話的長輩轉交給家中年輕人），但執行難度與拒訪率依然很高。

2. **涵蓋誤差（coverage error）：**現在許多在外租屋的大學生、上班族與小家庭，家裡根本沒有

裝設室內電話（稱為「唯手機族」）。在純市話抽樣框中，這群人被抽中的機率是 0。

由於原始收回的樣本結構明顯偏向高齡族群，若未經比例調整，調查結果就會過度反映長輩的意見，難以代表全體國民的真實想法。因此，事後的加權便成了不可或缺的校正步驟。

三、加權是怎麼運作的？從「多變量反覆加權」到「雙底冊合併」

為了讓樣本結構精準還原台灣 2,300 萬人的真實樣貌，現代台灣民調的加權通常包含兩個層次的精密計算：

1. **人口結構校正：多變量反覆加權（raking）**

民調機構會拿內政部最新的戶籍人口統計資料當作基準，透過統計軟體進行反覆迭代計算。如果真實母體中「20-29 歲女性」佔 7%，但這次電訪只抽到 3.5%，就需要把這群年輕女性受訪者的每一份問卷權重放大為 2 倍；反之，若「65 歲以上長輩」在樣本中佔了 40%，但是在實際人口僅占 20%，就需要把每位長輩的權重縮小為 0.5 倍。統計軟體會依據地區、性別、年齡、教育程度等多個維度進行交叉反覆調整，直到樣本的人口比例與內政部資料完全吻合為止。

2. **市話加手機雙底冊抽樣**

為了解決「唯手機族」漏接市話的問題，近年台灣主流民調機構紛紛導入「市話加手機雙底冊抽樣（dual-frame sampling）」。但這也帶來一個新問題：市話與手機是兩份不同的抽樣名單，且「家裡有市話、自己又有手機」的雙棲族群在兩邊都會被抽到，究竟該怎麼合併才合理？

臺大政治系洪永泰教授曾指出，若直接將市話與手機樣本按「1:1（各半）」或其他固定比例硬湊合併，在抽樣邏輯上是不嚴謹的。針對台灣國情，洪永泰教授與其研究團隊提出了一套更務實的「組合估計法（又稱聯立方程式法）」，透過「依電話使用族群結構進行事後分層加權」來合併資料：

A. **將全體民眾切分為三大電話使用族群**

訪員在開始正式訪問前，會詢問受訪者的電話持有狀態，藉此將樣本精準區分為三群：「唯市話族」（家裡只有市話、無手機）、「唯手機族」（個人只有手機、家裡無市話），以及「雙棲族」（市話與手機皆有使用）。

B. **依三大族群的母體估計比例進行「事後分層加權」併檔**

由於台灣缺乏官方公佈的「電話使用習慣人口普查」，洪教授團隊透過聯立方程式推算出台灣母體中這三類族群的占比（例如唯市話約占 1 成、雙棲族約占 6 成、唯手機族約占 3 成）。在合併市話與手機兩份資料檔時，則是依據這三群人在母體中的結構比例給予權重，並將兩邊抽到的「雙棲族」合併計算後調整權值，最後再與性別、年齡、地區等戶籍人口資料進行多變量加權，從而讓兩份不同來源的樣本無縫還原成單一且具代表性的全國母體。必須要說明的是，上述只是併檔的概念性說明，實際操作時母體結構的劃分、加權的順序、加權的次數、雙棲族的比例計算等仍有不同的處理方式，目前還沒有定論。

此外，洪永泰教授在實證研究中亦提醒，雖然雙底冊加權能大幅提升樣本涵蓋率，但「唯手機族」的拒訪率高且政治態度往往較極端，若併檔及加權的處理不當，反而可能放大估計誤差；因此雙底冊實際併檔與加權方式的差異，也會讓同一份民調原始數據產生不同的統計結果。

四、加權不是萬能藥：它的極限與代價

1. 加權倍數過高，會讓少數樣本掌握過大「話語權」

雖然加權是現代民調的標準配備，但它並不是可以無限制使用的魔法。在統計學上，雖然事前的「分層抽樣」能讓民調變準，但事後如果加權倍數拉得太高，反而會讓少數樣本掌握了過大的「話語權」，導致整份民調的抽樣誤差跟著放大。

舉例來說，如果某次純市話調查只抓到 10 個年輕人，卻硬是用加權把這 10 個人的意見放大 10 倍去代表全台年輕人，只要這 10 個人裡面有 2 個人想法比較極端，整份民調就會嚴重失真。因此，在實務運作上，許多民調機構除了花費較高成本執行「市話手機雙底冊」或「戶中抽樣」之外，有時也會以「配額」的概念進行「樣本配置」(例如設定特定年齡層或區域的最低回收樣本數)，來防止隨機的樣本結構過度傾斜；再配合事後的統計加權，讓最終統計數據既能維持一定的隨機基礎，又不至於因為加權倍數過高而失真。

2. 單一縣市的手機成本困境與「資料庫撥打」偏差

然而，當調查範圍從「全國」縮小到「單一縣市」(例如縣市長或區域立委選舉) 時，手機抽樣就會面臨一個致命的成本難題：台灣的手機號碼 (09 開頭) 是全國通用的，完全無法像市話 (02、04、07) 一樣從區碼分辨居住縣市。

如果某機構要做「嘉義市」或「新竹縣」的純手機隨機抽樣，電腦亂數撥通 10 通電話，可能只有 1 通是當地民眾，其餘 9 通都是外縣市民眾，必須禮貌掛斷。這種高達八、九成的篩選淘汰率，會讓單一縣市全手機民調的執行成本飆升到天價。

為了降低執行成本，部分民調機構在進行地方調查時，會改用既有的內部資料庫來撥打電話。這些名單通常來自過去累積的受訪者紀錄、商業行銷資料或特定會員庫；因為名單已經事前標記好居住縣市，所以撥打起來效率極高。然而，這類名單裡的人往往屬於較願意受訪或曾留下個資的特定群體，本身就帶有明顯的選擇偏差，並不具備隨機抽樣的代表性。由於事後加權只能調整性別、年齡、教育程度與地區等外在人口比例，無法改變抽樣來源本身不隨機的本質，因此即使經過加權計算，最終呈現的結果仍可能偏離當地選民的真實意向，這也是部分地方選舉民調與實際開票結果出現落差的原因之一。

結語

總結來說，「加權」是現代民調在預算考量與統計精確度之間，尋求平衡的必要工具。它能校正電訪中常見的年齡與接觸偏差，並協助整合市話與手機雙底冊的調查結果。然而，加權的效果仍取決於調查前抽樣的嚴謹程度，它能調整人口結構的比例，卻無法取代真正的隨機抽樣。理解這些基本原則，能讓我們在面對各類民調報告時，保有更冷靜且客觀的判斷力。

Ipsos 益普索市場研究

Ipsos是全球頂尖的專業市場調查研究機構，在全球九十個國家及地區設有分公司，在挖掘顧客經驗，執行品牌行銷調查，我們具有創新的思維與先進的科技。我們是世界頂尖企業執行長最信賴的企業顧問，我們也有熟悉各產業的專家協助測量、模擬和管理顧客與員工關係。

想了解更多資訊，請參考 https://www.ipsos.com/en-tw/solution/overview#category3

或洽詢我們的研究團隊

專案經理 Roger Kuo 郭銘忠

02 2701-7278 ext.207 Roger.Kuo @ipsos.com

Observer 副總監 Jill Huang 黃秋如

02 2701-7278 ext.216 Jill.Huang@ipsos.com