



一、應用統計、抽樣實務的樣本代表性

第二類，是「應用統計」與發生在抽樣實務過程的意義。

教室「統計學」常有許多「理想的假設前提」，當前許多統計學教科書上的習題，在實務上、真實社會中可能很難發生，其解答其實也是不盡完整的。（注¹）

「統計」是一種測量工具，研究者應正確的去使用它，達到我們的實務應用目的。但在實務界中，我們也發現常態性的把「統計名詞」當成儀式，並沒有正確使用統計工具解決問題的情形。（注²）

故本研究建議從4個方向來反映、分析實務上的「樣本代表性」：

1. 樣本是否具備隨機性/等機率性？
2. 樣本數影響的程度為何？
3. 抽樣方法影響的程度為何？
4. 樣本在「訪問實施」過程中，受到影響的程度為何？

(一)隨機性/等機率性與母群清冊

「推論」的基礎就是抽樣理論的「隨機性 Random」。

隨機性的中譯一詞非常難懂，吳統雄建議改譯為「等機率性」--亦即母群中每一個樣本都有相同的機會被抽中。

在本文中，除了已經流通的專有名詞（如簡單隨機）外，作為陳述均寫作「隨機/等機率性」，一則不違反習慣，再則強調科學中文化，以助讀者理解與明確的基礎認知。

而在抽樣實務上，就是要掌握完整的「母群清冊」，每一個樣本都要在清冊之中，才有被抽中的平均機會；不在清冊中的樣本比率愈高，代表性將愈低。

（注¹）看到這裡，有些專家大概要罵我竟敢批評教科書；客氣的人就在心中數落我大言不慚。當年，世界上絕大多數德高望重的智者，也都是這樣罵 Galileo，甚至要把他殺掉。如果你現在不贊成，請暫時把「聖經」放下來，走到第一線實際處理調查問題，實作很多、很多次以後，你也許將也會發現我將提出來的許多問題。對這些問題，我也會提出解決的建議。作為一個學習者，我不敢保證我的建議一定是對的；但作為一個真正的學習者，我知道自己不是在背書，而是獨立的、有興趣的發問與思考，而且對問題嘗試回應。

（注²）一個明顯的例子，就是各種媒體經常出現對選情的民意調查，在報導最後總有這麼一段話：「本調查樣本數為一〇六七人，使用電話簿抽樣，有效樣本數占百分之九十五，在百分之九十五的信賴區間下，誤差為正負百分之三。」這句話在近年來已經成為民調報導的樣版，以及教育讀者的習慣，這個傳播「不精確常識」的現象，令我們深以為憂。而在這些調查案後面，都有大群的教授、博士主持或顧問，卻也毫無知覺。



母群的範圍通常分為2類：小團體、或公眾。

小團體母群如：學校學生、公會會員...等，成員人數有限、資格隸屬明確，其母群清冊「完整性」高，「取得性」也高。

公眾母群如：大臺北市民、網路使用者...等，成員人數龐大、成員所在不明確，其母群清冊「完整性」與「取得性」都有挑戰。

母群清冊的形式也可分為兩類，一類是「實體母群清冊」，譬如文件型的電話簿、電腦資料庫...等；另一類是「虛擬母群清冊」，將在「[隨機/等機率性與建立母群清冊](#)」專章中詳細介紹。

(二)樣本數

上節指出，「[樣本分配](#)」的「樣本的平均數(\bar{x})」和「母群的平均數(μ)」的最大誤差(Maximum error of estimate)特稱為「標準誤」E，其定義公式為公式5.1：（注³）

公式5.1

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

其中，

$z_{\frac{\alpha}{2}}$ ：Z是指常態分配；有(1- α)的機率， \bar{x} 會在 $\mu \pm E$ 的範圍內。

σ ：母群的標準差（即母群變異數的平方根）

n：樣本數（有些書上寫 n-1，那是小樣本的情形；不過，對調查法而言，小樣本沒有意義。對大樣本而言， $n \approx (n-1)$ ，沒有差別。）

從此一公式將可看出：誤差與：

- Z值成正比；
- 母群標準差（亦即其變異數的開方）成正比。
- 樣本數的多寡成平方根反比；

而樣本代表性與誤差相反，亦即與：

- Z值成反比；
- 母群標準差（亦即其變異數的開方）成反比。
- 樣本數的多寡成平方根正比；

（注³）這項觀念稱為「中央極限定理」，詳見：

<http://tx.liberal.ntu.edu.tw/Jx/Methodology/Analy-TxStatisticsCanon.htm#central>



樣本數是研究者所能控制的。所以「樣本代表性」可以從樣本數的設計，來作預測或分析。

(三)抽樣方法

母群標準差我們不能預知，而是由「推算方法」的樣本變異數間接探求。就抽樣理論的目的而言，就是發展「抽樣方法」與「推算方法」。抽樣方法愈良好，在等量研究資源下，其推算式之變異數將愈小，亦即其樣本代表性愈好。

對以理論統計為取向的學者而言，其研究重點就是發展推算方法，將牽涉許多數字與公式。而對以應用為主要取向的學者而言，可能只需要知道某種抽樣方法可在何種情狀之下適用，大致可以知道這種方法對樣本代表性影響的優劣便足夠了。

(四)訪問實施

訪問結束以後，由於實務上不可能完全訪問到原始樣本，因此真正的樣本代表性極可能和設計階段發生差異，而須再考慮「受訪樣本」的狀況，與其對代表性的影響。

因此，就執执行程序而言，影響樣本代表性高低的因素，包括：

- 1、掌握母群清冊的程度。
- 2、決定適當的樣本數。
- 3、設計與選擇適當的抽樣方法，包括抽選樣本戶與抽選受訪者。
- 4、「受訪樣本」的狀況。但這部分是調查事後分析，我們將在「資料分析」章再討論。

二、隨機性/等機率性決策-建立母群清冊：樣本庫

「推論」的基礎就是抽樣理論的「隨機性 Random」。

隨機性的中譯一詞非常難懂，本研究者吳統雄建議改譯為「等機率性」--亦即母群中每一個樣本都有相同的機會被抽中。

而在抽樣實務上，就是要掌握完整的「母群清冊」，每一個樣本都要在清冊之中，才有被抽中的平均機會；不在清冊中的樣本比率愈高，代表性將愈低。

調查者必須「取得」「完整的」「母群清冊」。母群清冊可以是名單、地址、戶籍冊、電話簿、電子郵件地址…。樣本必須依據母群清冊抽出，母群清冊如果不完整，則不在清冊上的人就完全沒有被抽中的機會，從而發生了不隨機的問題。



(一)母群與清冊的種類

母群清冊的來源，可以分作「特定小團體」和「公眾」兩個方面來討論。

在調查特定小團體，譬如律師公會的會員或是某一家報社的記者時，母群清冊較易取得，完整性較高，隨機效果也較好。

而在調查一般公眾時，母群清冊便不易完整取得，便要檢查不完整部分的誤差，是否能夠容忍。

母群的範圍通常分為2類：小團體、或公眾。

小團體母群如：學校學生、公會會員...等，成員人數有限、資格隸屬明確，其母群清冊「完整性」高，「取得性」也高。

公眾母群如：大臺北市民、網路使用者...等，成員人數龐大、成員所在不明確，其母群清冊「完整性」與「取得性」都有挑戰。

母群清冊的形式也可分為兩類，一類是「實體母群清冊」，譬如文件型的電話簿、電腦資料庫...等；另一類是「虛擬母群清冊」。

1、實體母群清冊

實體清冊包括傳統文件型，以及資訊時代的電腦資料庫...等。常見的公眾實體清冊有以下幾種。

(1)地圖

美國於二次戰後興起的社會調查，最早採用的是造府訪問，由於沒有戶籍檔案制度，一般使用「區域抽樣法 Area Sampling」(Jessen [204])，亦即以地圖來抽樣，應用的母群清冊是地圖，抽樣單位是地圖上的家庭住宅。

二次戰前，城市鄉鎮建築更新慢，地圖完整性高。戰後快速都市化，市鎮發展變化大，衡諸地圖更新出版時間，已無法跟上即時完整性，也會遇到人口遷移或空戶的困擾，損傷抽樣的隨機/等機率效果，逐漸超過可以容忍的誤差範圍了。

● 母群清冊取得性

非大都會，較少人製作家戶地圖，不易取得。

故以地圖為母群清冊之實務已沉寂數十年，僅存在於古老文獻中。

(2)戶籍資料

擁有狹義戶籍資料的國家，全球其實不多。但廣義的戶籍資料包括各種身分登記證明，如美國的社會安全卡（工作證）、駕駛執照，我國的健保卡等。



美國曾經使用駕駛執照登記簿作為母群清冊，結果產生重大挫敗的經驗。本文將在「[樣本數決策](#)」專章中再詳細介紹，並說明不要混淆「隨機/等機率性」和樣本數的問題。

臺灣在國際上具備一套獨特的、歷史悠久的戶籍檔案，理論上可以提供完整的母群清冊，實施「戶口抽樣法」，且應有相當的效果。但是洪永泰發現，事實上當前的戶籍資料正確性相當低，已經無法代表真正的人口，並不適用作為母群清冊了。

● 母群清冊完整性

除非是極權國家，很難限制人民遷徙自由，也難要求人民遷徙一定要登記。

臺灣在國際上具備一套獨特的、歷史悠久的戶籍檔案，理論上可以提供完整的母群清冊，實施「戶口抽樣法」，且應有相當的效果。但是洪永泰發現，事實上當前的戶籍資料正確性相當低，已經相當不完整，無法代表真正的人口，並不適用作為母群清冊了。

而對個人有利益性的身分登記，如社會安全卡、駕駛執照，健保卡等，一般而言較為完整。

● 母群清冊取得性

在隱私權觀念興起後，各種身分資料都已受「個人資料保護」的限制，依法並不容易取得。

(3) 電話簿

在市內電話-即有線電話未百分之百普及前，有些人不能被電話連繫到，此清冊即不完整，便違反了隨機/等機率性的假設。

當電話普及率不高，或是沒有電話的人，如果具備某種共同特質，譬如貧窮、宗教信仰等，就會減低代表性。

Dillman [148]引述美國普查局的資料，指出美國的有線電話普及率於1975年達到90%。

「電話未能完全普及」是電訪的一個困難，但卻不是「致命傷」。Wolfe[375]指出當沒有電話的家庭比例不太大時，他們的人口特徵、意見取向和全體人口並無顯著不同。當電話普及率達到90%，已相當滿足隨機/等機率需求。其間的誤差是可以容忍的。

臺灣電話發展的情形也使樣本代表性的情形日益改善，於1980年後達到90%，在其後幾年甚至接近百分之百普及率。

因此，造成當時是電話調查準確度的黃金時期，本研究者也是在此時期在臺灣導入電話調查。

在社會環境的演變下，電話訪問在1980年後的30年間，成為較佳的選擇，因為它的隨機效果就與擁有電話的家庭比率成正比。

● 母群清冊完整性



電話簿一度是最佳母群清冊，唯2000年後陸續發生的2大社會行為趨勢，會使研究者對電話號碼掌握不全，有些樣本永不會被抽中，等同母群清冊不完整。

這2大社會行為趨勢是：電話簿不登記者增加，以及行動電話的衝擊。本文將在「[電話調查抽樣法](#)」專章中，再詳細說明。

● 母群清冊取得性

在市內電話普及之初，電信業者均提供電話簿，取得輕易。

唯電話技術數位化後，已罕見實體電話簿。電信業者的資料庫中，雖有完整清冊，依法也不容易取得。

在實務上，除了母群是極小、極集中的特定團體外，「實體母群清冊」經常不容易達到「完整」的狀態。所以「在實務上」，樣本代表性其實是指可推論「清冊所能夠對應的（部分）母群」，這是調查方法必然的限制。

2、虛擬母群清冊

由於實體母群清冊以上的各種困難，必須考慮如何建立虛擬母群清冊。

(1) 電話隨機撥號 (Random Digital Dialing, RDD) 技術

典型的虛擬母群清冊，就是「電話隨機撥號」(Random Digital Dialing, RDD) 技術的思考依據。

如在某些區域所有的家庭電話都是8位數，所以，所有的電話一定存在於包括從「00000001」號至「99999999」號的「虛擬架構-或想像的電話簿」中。（其他區域如電話為7位數等，均可類推。）

● 母群清冊完整性

理論上一定完整，但實務上發生新問題，如空號增加等，其挑戰與解決，本文將在「[電話調查抽樣法](#)」專章中，再詳細說明。

● 母群清冊取得性

虛擬母群清冊由研究者建立，沒有取得問題，但存在是否能夠建立的問題，如行動電話，就無法以相同邏輯建立。

(2) 發展中虛擬母群清冊

個人通訊行動化，是當前調查如何取得母群清冊最困難的問題。

不過，科技造成的挑戰，還是要由科技來解決。

網路大數據分析，可以提供一個補強的參考架構。

而數位電視機上盒、與「Google 地球」，有可能形成史上最有效益的清冊，協助發展未來的「行動調查」。