



一、樣本分析與代表性檢定

本文「[樣本代表性](#)」章指出，「樣本代表性」應有兩類意義：
第一類是「統計學、計量分析思想」的意義。
第二類，是「應用統計、抽樣實務」的意義。

就「統計學、計量分析思想」而言，樣本代表性的基礎就是「[中央極限定理](#)」，檢定方法是「顯著性考驗」，實則就是「事後樣本數檢定」。

實務上的「樣本代表性」，本研究者建議從4個方向分析：

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？

前3項，是在收集資料過程中已經決定，第4項就有待事後檢定。

統計理論基本上假設每一個樣本均可被調查完成，但是以公眾為對象的研究，幾乎不可能完成百分之百的訪問。統計學家已經證明「訪問完成率」對資料的統計與推論影響很大，因此必需定出一個對未完成率可以容忍界限。

依據統計原理，當母群很大，而不符隨機性的樣本很少時，後者可以忽略不計。

就這方面過去與當前的研究文獻來看，理論與實務的差距很大，本研究者將進行對分析與檢定標準的深入討論。

(一)檢定的指標

樣本代表性「檢定的指標」是什麼？它是一般調查者幾乎都沒有注意到的問題。

當我們作一個調查，觀察2個變項(Variable, 亦即訪問的問題)以上的時候，對一組樣本而言，每一個變項的差異都可能不一樣，精確的說，此時只有「變項代表性」沒有「樣本代表性」。

譬如，我們訪問一組樣本，人人樂意回答「性別」為何，但很少人願意回答「收入」多少。所以，這組樣本的「性別」代表性可能很高，但用這組樣本的「收入」來推論全體民眾的收入，代表性就可能不高。並不是「性別代表性高」，「樣本代表性」就高。

在調查報告中常看到的「完成率」、「成功率」、「有效率」...到底是什麼意義？彼此的異同為何？具體的操作定義為何？似乎仍然基於各研究者的自由心證。



一如Dillman [149], Kuiz [227]指出，「電訪問完成率」迄今尚無統一的定義，常用的電話調查完成率定義為：

$$\begin{aligned} \text{完成率} &= \text{完成訪問樣本} / (\text{總樣本} - \text{未曾接通電話的樣本}) \\ &\text{亦即：} \\ &= \text{完成訪問樣本} / (\text{回應樣本} + \text{拒絕樣本}) \end{aligned}$$

這項定義不僅過於簡化，而且容易引起語意混淆，無法清楚區別什麼是「完成」樣本。

第一、樣本的結果不僅於回應、拒絕、未接通。

Cochran [129]指出，一項調查可能還因以下原因而無法完成訪問：

- (1) 樣本戶中無人。
- (2) 沒有適當人選。
- (3) 樣本拒絕受訪。

Cochran 反映的還是傳統造府調查的問題，卻仍然無法涵蓋因為社會演變、新技術發展而產生實務上會面臨的各種問題。譬如：採用隨機尾數抽樣而產生的空號，家庭調查時接到工商電話，工商調查的對象已結束營業或行方不明，一個樣本戶卻有多個號碼，接到答錄機、傳真機或語音信箱等。

第二、逕行排除「未曾接通電話的樣本」會導致重大誤差。

如果受訪者為個人，許多研究已經證明，「未曾接通的樣本」-亦即不容易找到的受訪者，在人口特徵和社經地位上，和容易找到的受訪者往往有顯著的不同。把這種樣本從比率的分母剔除，實際上會降低樣本代表性。

借鏡上述經驗，如果受訪者為企業，亦應有相同考量。尤其在正常上班時間一再連繫都找不到的企業，其組織行為一定與正常企業不同，也許是一人公司；也許是公司已經實質結束，但電話仍然保留；也許是其他我們任何不知道的原因。但從研究「保守優先」的態度，當不能確認為不存樣本時，還是要歸為「未曾接通的樣本」。

「未曾接通的樣本」應包括：經由查證障礙臺、查號臺，仍無法證明為不存之樣本（不存樣本之定義詳後），而始終無人接聽，或無適當人選，超過時效仍然無法連繫者。沒有找到受訪者是調查者的責任，也是誤差的重大來源。

本研究者建議：找到受訪者稱為「連繫樣本」，「未曾接通的樣本」稱為「失連樣本」。

第三、把「拒絕」視為「未完成」是詮釋上的疏失。

「拒絕」表達的是：受訪者接到了訪員的電話，或是見到了訪員，但是知道是某些調查之後，不願意回答就掛斷了電話、關上了門。這個動作，事實上，已經表達了他的態度，我們必須認識「拒



絕」就是一種態度：消極的、沒興趣的、或是沒意見的態度。在全體受訪者的態度結構中，他們的代表性，是值得分析的。他們不應該被視為未完成、失敗的樣本。

吳統雄曾經指出：人類的本性，本來就不是對所有事情都有興趣的；民意結構中，占最大比例的就是「疏離者」。觀察拒絕的數字，可以看出積極民意，如何以少數引導多數；追蹤拒絕數字的消長，可以經由分析疏離者被動員的程度，預測危機事件的滋生與發展。

第四、迷失、有效、與拒絕之間的混淆

統計定義上，如果觀察「變項」出現「迷失值(missing value, 或譯「遺漏值」，但事實上不僅「遺漏」而已)」，稱為「無效(invalid)變項」否則即為「有效(valid)變項」。「有效、無效」是針對「變項」而言。但有些人分不清楚「變項」和「樣本」不同。一個樣本很有可能有很多觀察變項，其中有些可能有效、有些可能無效。所以「有效樣本」是一個極容易產生誤解的名詞。在統計程序上，是可以採用 list wise 方法，將含有任何無效變項的樣本剔除。但在調查實務上，如果是中等以上長度的問卷，一定有顯著比例的受訪者不會答完每一題，這樣處理後剩下的有效樣本必然偏低。我們經常看到一些調查報告，有很長的問卷、而很高的「有效樣本」比例，其實相當可疑。這種情形有可能是：對調查理論不了解、或故意造假、或管理不當以致訪員作弊。

實務上，產生「迷失值」的可能性有三：

(1) 沒有找到樣本。這其實不是「有效與否」的問題，而是更基本的「連繫與否」的問題。

(2) 人為疏忽：資料在登錄時遺漏、或登錄成不合理值，如在「性別」選項中出現「4」。在傳統手工調查時代，這種問題不小。但在調查資訊化的現代，只要系統設計嚴謹，這種問題可以降低到零。

(3) 「迷失」與「拒絕」混淆：許多調查沒有「拒絕」的選項，拒絕就變成「留白」、也就變成「迷失」。

如果調查的理論與實務嚴謹，調查資訊系統又得當，後面兩種情形其實是可以避免的。

第五、沒有考慮與處理樣本的更新（或替代）問題。

「抽出的樣本不可替代」是一個統計理論的鐵則，卻也是在實務上經常被忘記的事實。許多調查在很難找到受訪者，或被受訪者拒絕後，就逕行以其他樣本替代，導致真正的「未接通」與「拒絕」的數字混淆。

但在一種特殊的情形：由於母群清冊不明、或錯誤，包括：空號、在家庭調查中出現工商電話、在工商調查中的行號已結業、行方不明、電話故障經長期連繫依然不通（顯然違背正常企業經營型態）...等。



本研究者建議：發生上述特殊情形的樣本稱為「不存樣本」。
 樣本歸類為「存在」或「不存」必須經過查證確認，無法確認為不存者，均為「存在樣本」。如果發生「不存樣本」，研究者應可根據資源的有無，決定是否「更新(update)」不存樣本，以避免樣本數因不存樣本而減少。更新不存樣本無損樣本的代表性。
 「不存樣本」切不可與「失連樣本」混淆，如果逕行「替代(replace)」失連樣本，將有損樣本的代表性。

第六、未釐清「回應」與「拒絕」的界限
 受訪者回答所有的問題才算「回應」嗎？實務上，這樣的比率有限。回答幾題算回應？超過百分之50嗎？回答完應變項，沒有答復自變項算嗎？什麼又是「拒絕」？一開始就掛斷是拒絕嗎？答了1題、2題、3題...以後掛斷呢？

本研究者建議：調查者應該在問卷中自訂一個-或一組核心問題（項目）作為「回應標準」，有答的算「回應」，未答的計「拒絕」。這樣不致膨脹回應率，使得核心的資料低於表面數字；也不致於嚴苛，浪費了資料。

第七、沒有考量合格-「過濾問題」的狀況。
 在某些調查中，受訪者如果回答了核心問題，但是基於某些原因說謊，反而會誤導資料與分析結果。這時就要考慮實施「過濾問題」技術：如在選舉行為調查中，過濾受訪者行政區與選區的配合。
 本研究者建議：在實施「過濾問題」技術時，通過「過濾問題」的稱為「合格樣本」，未通過的則為「不合格樣本」。

本研究者依據以上的討論，建議將定義不清楚的「完成率、成功率」視為籠統的概念，容易混淆的「有效率」也暫時擱置。而將樣本分析分作具體精確的「4率」，其標準與操作定義如下：

不存樣本	由於母群清冊不明、或錯誤而產生的樣本
存在樣本	不能證明為不存之樣本，均為存在樣本
更新樣本	更新不存樣本
總樣本數	= 原始抽出樣本數 (+更新樣本數)
存在率	= 存在樣本數/總樣本數
連繫樣本	找到受訪者
失連樣本	找不到受訪者
連繫率	= 連繫樣本數/存在樣本數
回應樣本	回答核心問題
拒絕樣本	未回答核心問題
回應率	= 回應樣本數/存在樣本數
合格樣本	通過過濾問題



失格樣本	未通過過濾問題
合格率	= 合格樣本數/存在樣本數

以上的4率，連繫率、回應率、合格率，應與樣本代表性有關，存在率則與調查的成本效益有關。其意義如下：

1、連繫率

是影響樣本代表性最重要的指標。

連繫率高表示高比例的受訪者都經過最起碼的調查與測量；而未能連繫到的樣本，產生的誤差，只能歸咎於調查者。

2、回應率

表示獲得可量化資料的程度。同時，反映受訪者中積極、對問題參與性高的比例。

值得注意的是，有些調查是以「回應樣本數」作為報告數字的分母，在某些調查主題中，會產生膨脹數字的效果，甚至會將極少數的積極民意膨脹為優勢民意。

3、合格率

反映通過過濾問題，獲得較誠實、可信資料的程度。

這個指標是當有設計過濾問題時，才會產生。

4、存在率

當發生不存樣本時，一定會浪費連繫的成本，甚至會再增加更新樣本的成本，所以與調查的成本效益有關。存在樣本數，則應作為其他指標的分母。

存在率（注¹）也可以反映母群清冊的良窳。

（注¹）在應用統計上，在分析某一變項時，其中某一樣本的變項值如果不為「迷失值(missing data)」，則此樣本稱為「有效個案(valid case)」，有效個案占所有樣本的百分比亦稱為「有效率」。

「迷失值」的產生原因有二：第一、人工登錄資料發生錯誤；其次、問卷設計階段考慮不周全，發生了沒有預期的回應資料。如果在具備知識管理能力的資訊系統協助之下，「迷失值」將不再可能產生。故本研究建議放棄「有效率」舊有的內涵與定義，而採用本文的定義。