

抽樣統計調查實用指南



香港統計學會
Hong Kong Statistical Society

陳膺強 吳啟宏
何永煊 沈雪明



抽樣統計調查實用指南

何永煊

香港政府統計處

沈雪明

香港大學統計學系

吳啟宏

香港大學統計學系

陳膺強

香港中文大學社會學系



香港統計學會
Hong Kong Statistical Society

香港統計學會編輯委員會

主席：陳膺強

香港中文大學社會學系

秘書：王惠宜

香港政府統計處

委員：（排名先後以姓氏筆劃為序）

方惠焜

香港中文大學統計學系

吳啟宏

香港大學統計學系

梅克允

香港城市理工學院應用統計及運籌學系

陳禮忠

香港浸會學院校務處

蔡克敏

香港科技大學數學系

通訊地址：

香港薄扶林道

香港大學統計系轉交

香港統計學會

前 言

抽樣統計調查近來在香港非常普遍，而且應用範圍十分廣泛，預料這種趨勢亦將會持續下去。但應該注意，若統計調查實施不當，它的結果是沒有意義的。對於私人因商務理由而進行的調查，這或會導致不良的業務決策，但終究只會影響個別公司。但是一些要向社會大眾公佈結果的統計調查，卻可能影響到公眾意向和公共政策的制訂。負責這類統計調查的人員，應更嚴謹地工作，以盡其社會責任。

香港統計學會一向以不同的活動形式，提倡要正確地進行抽樣統計調查。學會在一九八二年和一九八四年出版了「統計調查是甚麼？」這本中英文小冊子，並在一九八七年向傳播媒介發出新聞稿，列出正確地進行統計調查的要點。學會並組織了幾個有關統計調查的演講及研討會。此外，學會更曾與香港大學校外課程部合作設立一個有關抽樣統計調查的課程。

香港統計學會編輯委員會認為，目前正需要一本較為詳細的小冊子，以提供有關怎樣正確進行抽樣統計調查的專業意見。於是邀請了學會會員何永煊先生（香港政府統計處）、吳啟宏博士（香港大學統計學系）、沈雪明博士（香港大學統計學系）和陳膺強博士（香港中文大學社會學系）撰寫。這書的英文版「A Practical Guide to Sample Surveys」已於一九九一年三月出版，現在向讀者介紹的是中文版。編輯委員會感謝四位作者的辛勞工作，並希望這小冊子對進行抽樣統計調查的人員和應用調查結果的人士有所幫助。

承蒙美國 SAS 軟件研究所贊助本小冊子的印刷費用，謹此致謝。

香港統計學會編輯委員會

序

以簡潔的方式介紹一門複雜的學科，是一項非常有挑戰性的工作。對香港統計學會付託的這個任務，我們誠惶誠恐地執行，並盡了最大的努力。在這中文版的編寫過程中，我們得到編輯委員會各委員及香港統計學界同寅的衷誠協助，謹此致謝。其中我們要特別感謝香港中文大學統計系方偉焜博士，他詳細閱讀了初稿，並提出了寶貴的意見。由於抽樣統計調查是一門博大精深的學科，編寫本書時難免掛一漏萬，希望讀者多多指正，以便在再版時加以改進。

何永煊 沈雪明
吳啟宏 陳膺強
一九九二年二月

抽樣統計調查實用指南

目 錄

第一章 統計調查的總體規劃	1		
1.1 肇定目標	1	3.10 避免採用「複合性問題」 (composite question)	16
1.2 對蒐集資料的質素要求	1	3.11 採用正確的「答案分項」 (response categories)	17
1.3 界定總體	2	3.12 良好的版式和編排的重要性	18
1.4 取得正確回覆的條件	2	3.13 應進行「問卷測試」 (pretest)	18
1.5 專業知識及一般資源需求	2		
第二章 樣本設計與抽選	4		
2.1 採用概率樣本	4	第四章 資料蒐集	21
2.2 考慮抽樣誤差	5	4.1 「試查」 (pilot survey)	21
2.3 決定樣本大小	6	4.2 訓練訪問員	21
2.4 一些常用的概率樣本設計	7	4.3 盡量減少無回應	22
2.5 採用合適抽樣架構	9	4.4 廉集無回應者的表面特質	23
2.6 不等抽選概率的適當加權	12	4.5 對回收資料的質素控制	23
		4.6 確保資料機密	24
第三章 問卷設計	13		
3.1 資料質量有賴良好的問卷設計	13	第五章 資料分析	25
3.2 問題應切合研究目標	13	5.1 多樣的分析方法	25
3.3 措詞一定要清楚、精確、不模稜兩可	14	5.2 基於簡單隨機樣本的估計	25
3.4 使用適當語言	14	5.3 基於較複雜抽樣設計的估計	28
3.5 採用別人設計的問卷時，要小心處理	14	5.4 注意分組後的樣本大小	35
3.6 記憶錯誤是答案不準確的重要原因	15	5.5 高回應率的重要性	35
3.7 必要時應設篩選性問題	15		
3.8 避免用「引導性問題」 (leading question)	15	第六章 報導統計調查結果	38
3.9 避免用「有傾向性的問題」 (loaded question)	16		
		參考文獻	40

第一章

統計調查的總體規劃

1.1 肇定目標

「統計調查」一詞，是用來描述一種以微觀著的方法。它透過向一些個體（即所謂樣本）蒐集資料，從而了解這些個體所屬的一個較大總體的情況。當我們需要某些資料時，便可能要進行統計調查。而首要的工作便是先界定調查的目標。通常這是主催調查者的工作。

清楚界定統計調查目標是非常重要的，一定要清晰、具體，不可模糊不清或模稜兩可。

1.2 對蒐集資料的質素要求

調查人員應保證蒐集回來資料的質素，使能切合預定的目標。

我們應了解到調查所得的統計資料，其價值不在於絕對準確；但有用的統計資料，必須要有一定的可靠程度，以滿足特定需求。對個別研究而言，因目標及方法不同，資料的可接受誤差幅度會有分別，故此並無衡量可靠性的通用準則。事實上在一些情況下，較為大的誤差幅度仍是可接受的。假設進行一個一般性的住屋情況調查，以了解房屋供應是否緊張。如果真實的住屋空置率很低，例如只有 1%，而調查結果顯示的比率為 2%，這可說是百分之一百的誤差，但卻不會造成顯著的害處。即使調查結果是 0.1% 或 3% 的空置率，都會得到相同的結論，即「住屋供應很緊張」。但假如調查的目標是要為未來建屋規劃，即興建多少單位提供資料，那麼便需要更準確的估計了。

要訂定一套適用於任何統計調查有關可靠性的準則，是不可能的事。只能從個別調查目的、資料用途、以及誤差大小帶來的後果而作出慎重考慮。上述因素同時影響到樣本大小、問卷設計、資料蒐集方法、以至工作人員的訓練指導等等問題。在資料分析及闡明

結果時，更應準確估計誤差的幅度。要得到準確程度較高的結果，通常需有較充裕的資源和工作時間。

1.3 界定總體

研究某一問題時，對象的全體便是「總體」。總體由稱為「個體」的基本單元，即分析單位組成。按某研究目標，分析單位可以是個人；但就另一些研究目標而言，分析單位可能是住戶、商業機構、學校或其他單位。

因界定總體的方法對調查結果的闡釋，有很大影響，所以調查總體必須就研究目標小心精確說明、清楚界定（如說明範圍、單位、時、地等）。

例如要進行一項調查，以了解屯門居民是否支持設立新的「屯門——元朗」巴士服務。當界定調查總體時，必須考慮有關的問題如(1)「怎樣才算是屯門居民？是否包括暫時住在屯門的居民？」(2)「訪問對象最少要有多大年紀？」等等。事實上當界定總體時，類似上述的問題便會出現，而「界定」的工作一般並不如表面看來那麼簡單容易。

通常可先按研究目標界定一個理想總體，即所謂「目標總體」。然後視客觀環境的限制，修訂為「調查總體」。這樣做的好處是可在界定過程中，察覺到不同限制的拘限程度和可能後果，然後適當地定下合乎實際需要的調查總體。

1.4 取得正確回覆的條件

統計調查只有在被訪者願意和能夠提供有關資料的情況下，才顯得有意義，故此必須盡可能取得被訪者的合作。

此外，負責調查者更應留意問卷內容是否過於敏感、有無侵犯個人私隱、與及問題是否太困難，以致被訪者即使願意作答，也根本沒法提供所問及的資料。

1.5 專業知識及一般資源需求

基於統計調查的技術性本質，進行這類工作時，必須具備專業

知識。從調查的最初階段開始，專業知識便不可或缺。因為當一個統計調查完成後才發現結果無效時，通常都難以補救。

要使一個構思完善的調查設計能順利落實，資源必須充裕，這包括了時間、金錢、人力等各方面。資訊通常是昂貴的；既然蒐集資訊要付出代價，那麼是否值得付出這代價，便要看所得到資訊的用處了。

有關本章所討論的內容可參閱 Moser and Kalton (Chapters 1, 2, 3) 及 Babbie (Chapters 3, 4)。

第二章

樣本設計與抽選

2.1 採用概率樣本

樣本設計是一項科學性工作，同時也是一門藝術，關乎怎樣從一個總體中抽取一部份進行調查。而首先要明瞭概率樣本與非概率樣本的分別。

進行概率抽樣時，總體中每個基本單元被抽中的概率必須為已知而且不等於零。這樣可避免主觀選取所引致的偏誤，同時可應用統計理論導出調查中「估計式」(estimator) 的特性。估計式是利用樣本資料來計算統計數值的數學公式。

非概率抽樣有多種不同方法：包括由自願者主動作答、偶遇式抽樣（但求方便，碰到誰便問誰），和判定抽樣（主觀地選擇認為有代表性的基本單元）。所有非概率抽樣法都有同一弱點——欠缺統計推論的理論架構。所以只能靠主觀審定去評估調查結果的有效性或誤差幅度，並無有效的統計方法可供應用。

故此，應盡可能採用概率抽樣法。利用概率樣本的資料，可科學化地作出有效的統計推論，估計總體的情況。

必須注意的是，有些表面看來可能相當客觀的抽樣法，其實似是而非。一個明顯的例子是定額抽樣法。以此法進行抽樣調查時，訪問員會按指示尋找及訪問各規定數目的「有代表性」樣本單元。至於樣本名額則是根據一些控制變項（如年齡，性別，地區等），把總體中單元分成類別，然後按比例分配。

例如 1986 年中期戶口統計有香港人口按區域及性別分佈資料。

區域	性別	人口數目	(%)
香港島	男	859,267	10.9
	女	856,593	10.9
九龍及 新九龍	男	1,195,663	22.2
	女	1,106,028	20.5
新界	男	966,245	17.9
	女	914,921	17.0
水上	男	21,289	0.4
	女	15,991	0.4

若要選取 1,000 成年人進行住戶意見調查，根據這些資料，便可算出各區的名額數目，如在港島區應選出 109 名男性及 109 名女性，在新界區應選出 179 名男性及 170 名女性（餘此類推）。

樣本名額決定後，即分發予個別訪問員，訪問員則按照名額所定，自行在住戶中尋找成年人進行資料蒐集工作（例如訪問 179 名住在新界的男性成人）。指定名額的目的在避免（或最低限度也限制了）訪問員在自由選取訪問對象時，主觀抽選帶來的偏差。但定額抽樣仍然不是概率抽樣，因為個別分類的單元並非用隨機方法抽選。雖然有名額限制，訪問員仍是按自己的意思抽選被訪者，故始終是偶遇式和主觀性的抽選方法。

2.2 考慮抽樣誤差

統計調查誤差的來源是多方面的，而從樣本得來的統計數值，都有抽樣誤差。抽樣誤差的存在，是因為每一次調查得來的估值，只不過是按同一抽樣法，抽出大小一樣但並不相同的樣本，進行調查所能得到的千萬個估值中的一個。

假定要採用從樣本中觀察到的比率 (p) 以估計總體中具有某些特質的個體的實際比率 (P)。

與此有關的一個重要問題是——在相同的設計下不斷地抽取相同大小的樣本，哪些 p 值可能會出現？而不同的 p 值出現的概率又怎樣？整列可能出現的 p 值，與它們出現的概率相對應，便是 p 值的「樣本分佈」 (sampling distribution)。樣本分佈說明 p 值隨樣本設計而產生的「隨機變動」 (random fluctuation)。這些不同 p 值的變動程度，叫做「變異性」 (variability)，可用「標準誤差」 (standard error) 來量度。標準誤差便是 p 值的樣本分佈的「標準差」 (standard deviation) 的一個估計，是一個統計分佈離散度的量度。

2.3 決定樣本大小

調查時所需的樣本大小，要視乎抽樣方法和所要求的估值精確度而定（即可接受的誤差幅度或抽樣誤差）。這可按一定的統計方法計算出來。下面的例子說明怎樣在實際情況下決定樣本大小。

假設要用面對面訪問方式進行調查，以估計一個有 15,000 成年人的總體中，百分之幾的人將會使用一個正在計劃興建的圖書館。要決定樣本的適當大小，必先要定下所要求的估值的精確度。這工作並不簡單，且開始時常會把要求定得過高。假定最初要求估值距離總體 P 值不超過兩個百分點 (0.02) 的機會為 95 %。換言之置信區間 (confidence interval) 為樣本值 p 加、減 0.02。即要求：

$$1.96 \times \text{s.e.}(p) = 0.02 \quad (2.1)$$

其中 p 為樣本百分比， $\text{s.e.}(p)$ 為 p 值的標準誤差（見 §5-2）。

假如採用簡單隨機抽樣法 (SRS)，並暫不考慮簡稱 fpc 的「有限總體修正因子」 (finite population correction factor，見公式 (5.1) 後的說明)，從基本統計理論可知：

$$[\text{s.e.}(p)]^2 = PQ/n' \quad (2.2)$$

P 為總體中的比例

Q 等於 $(1 - P)$

n' 乃所需樣本大小的初步估計

數學上，可知 $P = Q = 0.5$ 時， PQ 的值最大，故一個較保守的方法是從最可能的 P 值中，取最接近 0.5 者，作為 P 值的估計，以供計算 n' 。如設 P 值為 0.4，則根據公式 (2.1) 及 (2.2) 可知

$$n' = 2305$$

如覺得 P 值無從估計，則可於計算 n' 時設 $P = 0.5$

如相對於總體而言，樣本相當小（因有一個很大的總體），計算時可略去 fpc，而 n' 即為所要求的樣本大小。不過相對於 $N = 15,000$ 的總體而言， $n' = 2305$ 並不算細小，所以必須考慮到 fpc。據統計理論，可用下式求得樣本大小 n 。

$$n = Nn'/(N + n'). \quad (2.3)$$

以式 2.3 可求得 $n = 1,998$ 。

在計算樣本大小時，另一個要考慮的便是可能出現的「無回應」 (non-response) 問題。假定回應率可達 75%（一般可按以往調查結果估計），則原定的樣本中 1,998 名成年人應更改為 $1,998/0.75 = 2,664$ 。這樣的調整並不能解決無回應偏差 (non-response bias) 的問題，這在下面將再行討論，但卻可提供足夠的樣本單位數目，滿足對 p 值精確度的要求。

到了這階段，研究人員可能會重新考慮開始時定下的精確度要求，看看有無放鬆一點的可能。因為精確度並非唯一的不可改變的要求，採用多大的樣本同時亦受到調查成本的限制，一般要在成本與精確度間找尋折衷辦法。

樣本大小的選擇要視乎對一系列數值的預計；例如上例的總體中有多少比例 (P) 成年人回答說會使用圖書館，和無回應率有多高等等。預計不準確會導致估值的精確度與原來定下的不同，不過估值本身仍是「總體參數值」 (population parameter) 的一個合理估計。

如果應用較複雜的抽樣法，那麼要決定樣本的大小便要用其它標準誤差估計公式（如 §5.3 所示）。

2.4 一些常用的概率樣本設計

「簡單隨機抽樣法」 (simple random sampling) 基本只是以抽簽方式抽取所需數目的單元，但其中也包括了一些不盡相同的抽

選方式。

另一個常用的方法是「等距抽樣法」(systematic sampling)，即常見的每 k 個單位抽取一個的方法。此法先定一「隨機始點」(random start)，例如一列單元中的第 m 個 ($m \leq k$)，然後順序抽出第 $m + k$ 、 $m + 2k$ 、……各單元構成樣本。

很多時有關調查總體中基本單元的若干輔助性資料是已知的。這些輔助性資料可用來改進樣本設計；如採用「分層」(stratification) 的方法。分層的意思是據已有資料把總體分為若干個不重覆的次總體——「層」(stratum)。分層時是根據一些與研究變項有關的特質，把單元歸入各個「同質性」(homogeneity) 較高的層，然後分別從各層抽取獨立的樣本。分層抽樣的好處是各層樣本的大小都受到控制，而非在抽樣過程中由隨機決定。同時調查人員可作出更精確估計（統計理論上確是如此），而且不同的層更可採用不同的抽樣方法。

在很多抽樣設計中，都把總體的基本單元分成多個組別。上面提到的分層抽樣法便是把這些組別作為「層」處理，從每組中抽取樣本。另外一種抽樣法把組別作為「羣」(cluster)，在所有羣中抽出若干羣作為樣本進行調查，（但起碼要抽取兩個羣，否則抽樣誤差便沒法計算）。當樣本包括了被抽出的羣中所有的基本單元時，便稱為「整羣抽樣法」(cluster sampling)。但如在被抽中的羣中，再抽出部份基本單元進行研究，則稱為「兩段抽樣法」(two-stage sampling)。在較複雜的抽樣設計中，更常會採用一系列等級不同的羣。例如一項對香港學生進行的調查，可能先抽出一個以學校為單元的樣本，再從每間被抽中的學校抽出若干班，最後從抽出的班中抽取學生樣本。這便是「多段抽樣法」(multi-stage sampling)

簡單隨機抽樣法一般被視為有用的基準，用以與較複雜的樣本設計作比較。假設最後估計的精確度要達到某一特定水平時，需要一個樣本大小為 n 的簡單隨機樣本。如調查負責人因某些原因採用整羣抽樣法，則可以預期他將會需要一個比 n 大的樣本。另一方面，他如採用適當的分層抽樣法，則可能一個小於 n 的樣本便

足夠。在各種不同因素中，抽樣方法的選擇，對樣本大小有重要的影響，而後者則對最終估計的精確度有決定性作用。

2.5 採用合適抽樣架構

無論採用那一種概率抽樣法，最基本的要求是先要有一個「抽樣架構」(sampling frame)，以便從中抽選樣本。故此抽樣的首要工作是取得有關總體的合適抽樣架構。

當具備總體所有基本單元的清單時，此清單便可作為抽樣架構。但如沒有清單時，則必須有一個可以識別和找出基本單元的方法。例如進行「區域抽樣」(area sampling) 時，總體中每一基本單元必屬某一特定地理區域——「地區」(area segment)，此地區一定要有清楚劃分的天然或人為邊界。個人或住戶均透過住址（如有多個住址，則以常居住址為準），與地區產生聯繫；所以每個單元必隸屬單一地區。故當一個以地區為單位的樣本選出後，訪問員可檢核所有被抽出地區內的適當單位（例如住戶），並編訂清單，跟着便可決定抽選部份單位或以全部單位為訪問對象。

在抽樣設計中，樣本架構可說是一個主要部份。它既提供界定和找出總體中基本單元的途徑，同時更常涵括了相當數量的補充資料，有助分層或分羣的工作。

理想的抽樣架構，只登錄總體中每個基本單元一次，不會重覆，而其他單位並不登錄。但實際情況下此理想甚難達到，故研究員應留意每個架構的缺點。通常有四類潛在的架構問題，此乃「遺漏單元」、「羣」、「架構外單元」、「空白」及「重覆登錄」。茲分別討論如後。

(a) 「遺漏單元」(missing elements) :

指總體中部份基本單元未被包括在架構內。這是因為抽樣架構「不充足」(inadequate) 或不完整 (incomplete) 的緣故。前者是指抽樣架構與目標總體不完全吻合，故此已知必有某些單元不被包括在內。而後者是指雖然有理論抽樣架構與目標總體互相吻合，但實際上仍漏去一些基本單元。兩者的區別在實際應用上相當重要，因不充足的情況一般較易看出來。例如在一個以某校學生為對象的

調查中，不把兼讀學生（其實亦是目標總體的一部份）包括在內，那名單便是不充足。若學生名單稍過時，一些新來的學生可能沒有包括在內，那就是不完整。

遺漏單元是架構問題中最嚴重的，因為除非有補救的方法，否則那些單元便永不會有被抽選的機會，樣本便沒法代表整個目標總體。有時可以重新界定調查總體，剔除所遺漏的單元，以避開上述問題。但這是個不完善的解決方法，只有在所剔除的一羣僅佔總體的一小部份，且對調查目標影響甚微，又無其他簡易可行的解決方法時，才會採用。

一個較可取的解決辦法是尋找一個補充架構以包括所有遺漏的單元。例如在上述學生調查可加編兼讀學生名單及新入讀學生名單。但是這方法可能會帶來重覆問題，因為有些學生可能在不同名單上同時出現。不過這問題較輕，可用下述有關重覆登錄的解決方法去處理。

當無法找到遺漏單元的補充架構時，便須採用某些「連繫程序」(linking procedure)去解決。連繫程序的目標是在把遺漏單元和加編名單上的單位清楚地聯結起來。當這名單上單位被抽選中時，它本身所包括的基本單元和與它連繫的遺漏單元便一同被作為一個羣處理。故此連繫程序會帶來羣的問題，這可用下面討論到羣時所介紹的處理方法解決。

例如在上述學生調查中，樣本架構是原來每班學生按姓名字母排列的名單。一個可行的連繫法是，名單內每個名字除了代表該學生外，還代表該生與名單內下一位學生之間所有被遺漏的學生。為了不忽略位於名單上第一個名字之前的遺漏學生，整個抽樣名單可視為一環形名單處理。如此則所有應在最後一個名字之後或首名字之前，但被遺漏的學生，均連繫於名單上最後一位學生。這種連繫方式是「半開放區間」(half-open interval)的一個例子。這種連繫方法也常用於居住單位的抽樣。

(b) 「羣」 (cluster) :

有時清單上一個單位包括了整羣的基本單元，而非單一的基本單元。上面討論到的連繫程序有時便會產生羣。此外其他的情況亦會產生羣。例如當要抽取一個以個人為基本單元或住戶為基本單元的樣本，但抽樣架構是一個以居住單位構成的清單時，便會產生羣，每羣內有多個個人或住戶。

在這類情況下，一個解決辦法是把屬於抽中的羣內所有基本單元都包括在樣本裏。這時每個基本單元被抽中的機會，即等於該單元所屬的羣被抽中的機會。如名單上的單位（羣）是以等概率方式抽選，則基本單元被抽中的概率亦均等。不過這種「全取」的解決方式在原定樣本大小規限下，會導致有較大的抽樣誤差，或是因樣本大小比原來設計的大，令工作量的預算失準。

(c) 「架構外單元」 (foreign element) 和「空白」 (blank) :

架構外單元是指清單上的一些基本單元，並非是這次調查的對象；例如在一個受薪者的調查時，失業者便屬架構外單元。空白的意思是指清單上的某些基本單元已不再在總體中存在；例如地址清單上未剔除的已拆卸居住單住。

解決空白及架構外單元的方法很簡易，若抽選中時，只須設想該基本單元並不實際存在即可。抽選概率不會受影響，但這樣的樣本會比原來抽出的小一點（因抽樣時把上述的單位一起抽出而事後又省略去）。所以在決定「抽樣比」(sampling fraction)以期得到某一數目要求的樣本時，這因素應予以考慮。

一種常見的錯誤做法是：當空白或架構外單元被選中時，代之以清單上隨後出現的一個單元。這些有可能被間接選中的單元便有較大的被選中概率，故此不應採用這種做法。

(d) 「重覆登錄」 (duplicate listing) :

某些總體中的基本單元會被重覆登錄在抽樣清單上。當抽樣架構由幾份清單合併而成，而某些基本單元又同時出現在不同的清單上時，便會有這種情況。所產生的問題是被重覆登錄的基本單元有

較高的被抽中概率。而這概率的大小視該單元被重覆登錄多少次而定。

一個解決方法是從抽樣架構中把所有的重覆剔除，使每一基本單元只登錄一次。不過這做法耗費相當大。另一解決方法是採用「單一識別法」(unique identification)，把個別基本單元與它的所有登錄之中的一個清楚地聯結起來（例如第一個出現的登錄或最先存在的一個登錄），而其他的登錄作空白登錄處理。假如抽樣架構的組織，或所具備的資料不容許直接採用單一識別法，則仍可在實地訪問時再運用這方法去解決重覆登錄的問題。但因相當部份的調查費用已用於接觸被訪者的工作上，故於訪問後才剔除一些被抽選中的基本單元而作爲空白處理是不符合經濟原則的。另外一個方法是採納所有被選中的基本單元，然後用「加權」(weighting)的方法去調整不均等的抽選概率。在這個情況下，必須在實地訪問時找出個別樣本中基本單元有無或者有多少個重覆的登錄（可在訪問時向被訪者詢問適當問題），作爲加權的準則。

2.6 不等抽選概率的適當加權

要做好統計調查，並不須刻意採用等概率抽樣設計方法 (EPSEM，即所有個別被訪者都有相同抽選概率）。當採用不等概率抽選樣本時，不同被訪者的抽選概率必須爲已知，以便找出適當的加權方法進行分析，從而得出無偏差的總體估計。

所謂加權的意思，是在資料處理和分析的過程中，對樣本中不同的基本單元，配以不同的相對重要性，即所謂「權數」(weight)。一般而言，抽選概率較高的基本單元，會配以比例上較小的權數。

此章討論的內容可參考 Moser and Kalton (Chapters 4, 5, 6, 7) 及 Babbie (Chapters 5, 6)。

第三章

問卷設計

3.1 資料質量有賴良好的問卷設計

統計調查常以訪問或郵寄問卷方式蒐集資料；間中亦會採用觀察等其他方法（可參考 Moser and Kalton, Chapters 10, 11, 12）。「問卷」是把問題組織起來的一種表格，用以記錄有關被訪者的資料。

當以提問的方式蒐集資料時，是假設所得到的答案是可靠和有效的。不過答案的可靠性和有效性常常受到被訪者提供正確答案的能力和意願所左右。而被訪者對問題字句的了解，有無所需的資料，以及訪問環境等都會影響到作答的能力。一般人對涉及個人不愉快經歷，敏感性或尷尬問題，都會不大願意作答。特別是在不適當的訪問環境，或對訪問的保密性有所懷疑時，困難更大。

一份設計良好的問卷，是資料質素的主要保證。當設計蒐集意見的問題時，應留意答案每易受到措詞、問題重點以及結構等影響。這些問題屬主觀性，要由作答者憑主觀作答，因此與客觀性問題不同，一般難有外在準則可供判斷其有效性。故此問卷設計必須嚴謹，下列是一些設計問卷時應注意的事項（可參考 Moser and Kalton, Chapter 13）。

3.2 問題應切合研究目標

問卷涉及範圍不要過大，不應嘗試蒐集指定範圍以外的資料。問卷亦不宜過長，因爲問題越多時，費用亦會較高。而且訪問時間太長，會使被訪者覺得不耐煩，影響到答案的可靠性。故此應盡量避免加入一些並非絕對必要的問題。

3.3 措詞一定要清楚、精確、不模棱兩可

用詞過於簡略或含糊是導致答案不準確或無效的主要原因。

例：「請問你有多大年紀？」

—回答者可能以農曆或是以公曆計算，又可能以足歲或虛齡回答問題。

—可改為問出生日期，或說明以何種方法計算年齡。

「你是否不認為甲方案不是解決某問題的好方法？」

—雙負面意思問法，會使被訪者混亂。

—應改問：「你認為甲方案是不是解決某問題的好方法？」

「去過你有見過輔導主任嗎？」

—當被訪者回答「有」時，可能是指：

- (a) 見到過（眼睛看到），
- (b) 在某場合遇見他，或
- (c) 接受他的輔導。

—應改問「去過你有接受輔導主任的輔導嗎？」。（假設這是提問的目的）

此外，問題設計者要肯定措詞對所有的被訪者而言，都有相同意義。

3.4 使用適當語言

採用適當語言，包括在適當時採用方言，俚語和行內常用語，可保證不同類別的被訪者都能明白問卷中問題，提高答案的準確度。

3.5 採用別人設計的問卷時，應小心處理

當採用本來是外語問卷，或採納個別問題 / 尺度時，要小心考慮問題的內容和提問方式，看是否適用於本地社會文化，必要時應先作適當修改。至於問卷的翻譯工作更不能隨便。當問卷從甲語文翻譯成乙語文後，為檢定翻譯工作做得是否妥善，一個可行的方法是請另外的翻譯人員再從乙語文翻譯回甲語文，比較兩

份甲語文的問卷有沒有明顯差別，特別是文義方面的不同。如有，便要改善及重做翻譯的工作。

3.6 記憶錯誤是答案不準確的重要原因

在一段相當長時間前所發生的事，一般人都難記憶清楚。故此不宜提問涉及已過去相當時日，或在一段相當長時間內（如一年內）所發生的事情。假如無法避免這類問題，應請被訪者盡可能翻查紀錄然後作答。

3.7 必要時應設篩選性問題

部份問題之前宜加設篩選性問題（過濾性問題）。

例：「課餘兼職對你的學業成績有沒有影響？」

—當被訪者回答說「沒有」時，可能是指

- (a) 沒有兼職，或
- (b) 沒有影響。

—故應先問被訪者一問題，那便是

「你有沒有課餘兼職？」然後只向答「有」的被訪者提問上述問題。

3.8 避免用「引導性問題」(leading question)

有些問題在內容、結構或措詞方面會引導被訪者提供某類答案。

例：「你不會認為……，是嗎？」

—通常會帶來否定，即同意「不會認為」的答案。

「你是否同意政府有……的責任？」

—通常會帶來肯定的答案。

「你在家中的晚膳時，會看甚麼電視節目？（例如 A、B 等節目）」

—被訪者常會以問題中所提及的節目名稱作答。

3.9 避免用「有傾向性問題」 (loaded question)

通常有許多不同的方法，在問題內加入一些壓力，令作答的人傾向於某一立場，提供某一類答案。

例：「為了更好的……服務，你是否同意合理的加價？」

—合理一詞，有增加肯定答案的作用。

「子曰：『……』，你同意嗎？」

—通常受人尊重或著名人仕的名字，會使人傾向於同意一句陳述或一個意念。

「衆所週知……」

「按照法律……」

「……由來已久……」

—引述社會期望、現況、或典型事例等等，都會較易被人接受。

故此在提問時，應避免採用上述各種提問方法。

有時個別問題本身只是一句簡單而無偏倚的陳述，不過在所提供的答案選擇中，可能出現複雜程度不一的項目。這樣的問題便可能有傾向性，因為被訪者會傾向於選擇較易明白的答案項目。

3.10 避免採用「複合性問題」 (composite question)

問卷中不應有複合性問題（或一題兩問——double barrelled），此類問題常把兩個或更多問題合併在一起提問。

例：「你會否在一個月之內辭去現職並另找一份新工作？」

—如被訪者計劃在一個月內辭去現職但不打算找新工作，便無法回答「是」或「否」。

「你對香港的安定繁榮有信心嗎？」

—一題兩問，非「單維性」 (unidimensional)。

—應分成兩個問題提問，一問安定，另一問繁榮。

3.11 採用正確「答案項目」 (response categories)

一般在大規模的調查研究中，問卷中問題多採用封閉式，即預設答案項目供被訪者選擇。因相對而言，這比開放式問題在答問、編碼及分析上較有效率。

預設答案項目應該「盡列」 (exhaustive) 和「彼此不重覆」 (mutually exclusive)。而意見或態度問題的答案項目，更宜是對稱和兼有中間或中性類別。

例：「你住所的通風情況怎樣？」

- 很好
- 好
- 普通
- 差

—不對稱，正面的答案較負面的多。

—應改成：

- 很好
- 好
- 普通
- 差
- 很差

「你認為某人在……方面的表現怎樣？」

- 優異
- 良好
- 滿意
- 尚可
- 差劣

—非單序向 (monotonic)。因「良好」不一定超越「滿意」。

—應改成：

- 極佳
- 甚佳
- 普通
- 甚差
- 極差

「你滿意目前的工作嗎？」

- 極滿意
- 滿意，但感到有壓力
- 不滿意，但沒有壓力
- 全不滿意

一答案中把兩項尺度（滿意程度、壓力程度）合併了。

一應分成兩個問題，一題問工作壓力，另一題問滿意程度，後者更應說明是指那一方面，如時間、薪金等等。

很多時候答案項目中還得加入「不適用」和「無答案」（指被訪者不願作答）等項目。此外還得視乎問題性質適當地加入「無所謂」、「不知道」、「無意見」、「未決定」或「視情況而定」等項目。當決定那一項目應列為某一問題的答案時，問卷設計者應清楚考慮該問題的本義。

3.12 良好的版式及編排的重要性

問卷中的問題應按內容分成不同部份，加上編號並適當地排列。雖然並無一般通用的先後排列準則，但值得注意的是，當無可避免地要問一些敏感性或尷尬問題時，最好還是把問題放在問卷的末段。美觀的印刷，在適當地方加上說明（見附例甲）及清楚的指示如那些問題應跳問等（見附例乙），都可幫助減少提問和作答的錯誤。

3.13 應進行「問卷測試」(pretest)

正式進行訪問之前，問卷必須經過測試。這可幫助找出問卷是否符合調查所需，問題的措詞是否恰當，預設答案項目是否配合實際情況需要，次序編排是否合理，以及在訪問的場合，是否適宜向研究對象提出問卷中的問題等。

本章有關內容亦可參考 Babbie (Chapter 7)。

附例甲：括號內附加說明、指示及交替性字句

14. (只問有子女未入學或不送往托兒所、幼稚園之被訪者。)

通常你的子女每日有幾多小時不在家（屋企）內的呢？

_____小時

- 不知道
- 無答案
- 不適用

15. 在你工作中的機構（假如此機構有分支）則此題單指被訪者所在之分店、分行，或總行）一共有幾多人工作呢？

_____人

- 不知道
- 無答案
- 不適用

附例乙：跳問問題的流程指示

17. 你有没有全職工作？

有
 沒有

→ 17a. 為甚麼你沒有全職工作呢？

處理家務
 在學
 有其他經濟來源
 已退休
 找不到全職工作
 其他（請說明）_____

→ (轉問第七頁第 21 題)

18. 你從事那一行業的工作？

農業、林業、捕獵及漁業、採礦及採石
 製造業——紡織及成衣
 建造業
 批發及零售業、酒樓餐廳及酒店
 交通、倉庫及傳播事業
 財經、保險、地產及商業服務
 服務行業
 其他

第四章

資料蒐集

4.1 「試查」(pilot survey)

資料蒐集可用上門訪問、電話訪問、自填問卷或郵寄問卷等方式。選擇何種方式，則必須小心考慮被訪者的合作意願、所探討問題的複雜程度和其他有關因素。

有關實訪工作的各種細節，諸如分配訪問員工作量及安排交通等都要小心策劃，以保證工作能有效地順利進行。

實訪工作的程序應在正式進行前予以測試。事先測試問卷和實訪程序有無不妥是保證日後工作能順利進行的唯一辦法。這在採用新的調查方法或一份新設計的問卷時更形重要。要預計所有不同問題和工作程序可能帶來的誤會或偏差效果極為困難。所以一個完善的調查計劃必須包括「試查」。試查的觀測結果和經驗，可作為調整計劃的根據。

4.2 訓練訪問員

有關統計調查所採用的概念和定義，應小心向訪問員解說。在他們正式投入工作之前，適當的訓練更不可缺少。

訓練可用自學或課堂訓練形式，亦可二者同時進行。訓練應強調良好訪問技巧；這包括如何作初步接觸，如何進行合乎專業水平對答，如何避免影響或偏導回覆等等。訓練一般包括實習訪問，使訪問員熟習在調查時可能面對的各種不同處境。有關材料應事先準備好並分發予訪問員。應包括問卷的樣版、參考手冊、樣本單位的識別和定位的資料，以及會向被訪者展示的紙卡或圖片等。

訪問員要確保被訪者明白問卷中的問題。只在必要時才作較深入探索，且不可過量，以免被訪者受到不適當影響。

4.3 盡量減少無回應

調查應包括採用一系列措施，嘗試盡量減低拒答者的數目。進行訪問之前，調查負責人宜先致函樣本中單位，說明統計調查目的和將會派訪問員造訪。進行大規模的統計調查時，廣泛宣傳工作更為重要。調查如採用訪問方式，應小心訓練訪問員如何避免拒答。例如在首次訪問時發覺時間不合適（如住戶正在用膳），那訪問員便應在較適當的時間再次造訪。

若能說明調查的應用價值，將較容易取得被抽中的樣本單位合作。表明調查是由有聲譽的機構主辦，則幫助更大。主辦機構的良好聲譽在郵寄問卷調查中，效用特別顯著。統計調查的保密性，包括用不記名方式記錄資料，通常可消除被訪者對其所提供資料的用途可能產生的疑懼。一般問卷都是以較簡單和不會引起疑慮的問題開始，以免被訪者因面對尷尬問題而立即中止答問。

在訪問時，被訪者如「不在」，訪問員都會「再訪」(callback)。訪問員如無法接觸到被訪者，通常都會按指示進行三至四次的再訪，每次再訪應在不同的時間（白天或黃昏）進行。如訪問員剛在鄰近地區工作，可進行更多再訪。當樣本指定的被訪者不在時，訪問員可能的話應透過其家人預約訪問時間，這可增加成功訪問的機會。

進行電話訪問時，再訪便容易得多，故再訪的次數亦常較上門訪問時多許多。

用郵寄問卷方式進行調查時，與再訪工作相若的便是「跟進工作」(follow-up)。例如再寄出函件鼓勵沒有回答者作答。此外亦可採用混合的方法，即是當被訪者沒有利用郵遞回覆時，派出訪問員上門訪問。

一般而言，調查工作者都應盡可能減低無法接觸者和拒答者的數目。總而言之，足夠的宣傳工夫，明確介紹主辦機構及清楚證明訪問員身份，事先知會樣本中被訪者等等措施都會有幫助。此外更應妥善安排再訪及跟進工作，並勸說不願合作的被訪者參與。

4.4 蒐集無回應者的表面特質

雖然經過工作人員的不斷努力，但無回應者的數目有時仍不容忽視。如有這種情形，便需以適當方法幫助闡釋調查的結果。在這方面來說，能夠取得一些無回應者的表面特質資料是有用處的。

其中一種調整無回應的方法，是需要研究者掌握所有樣本單位（不論有無作答）的某些共有資料。例如某樣本按基本單元位處於農村、市郊或城市而劃分為不同的地理區域。如這是個等概率樣本，則調整方法是按不同區域的無回應率的高低而配以不同的權數；即以 $n(h)/r(h)$ 加權於 h 區域的作答者。其中 $n(h)$ 是樣本大小， $r(h)$ 是已完成調查的樣本單位數目。這樣的調整使到各個區域已完成調查單位的樣本分佈與整體樣本分佈吻合。這方法是利用每區的作答者透過加權步驟以代表無回應者。

4.5 對回收資料的質素控制

調查的進度，必須有適當的控制。過於緊逼的工作日程常導致實訪工作的質素下降。控制個別訪問員的質素亦同樣重要。這可通過多個方式進行。常見的是由高級或督導人員再訪或覆核個別訪問員所負責的部份樣本單位。同時經過辦公室校訂程序，查核資料有無遺漏或明顯錯誤。

當訪問完成，而問卷亦已填答完畢後，問卷必須經過人手處理才能把資料輸入電腦。這牽涉到把未預先編碼的問卷項目逐條編碼的文書工作。例如「職業」、「行業」的編類便是典型的繁複編碼工作，需小心控制質素。開放式問題的編碼亦如是。

抄錄、編碼和鍵盤輸入等工序均會有人為錯誤出現，故必須部份或全部覆核，嚴格控制質素。當資料已輸入電腦並存檔後，即可進行電腦的校訂工作，找出資料中矛盾和錯誤的條項（例如一個年僅六歲的人有祖父的身份等），並以「設算法」(imputation)，按已定的規則，更改此類條項。

4.6 確保資料機密

有聲譽的調查研究組織都很關注被訪者所提供的資料的私隱性。訪問員應向被訪者清楚解釋調查並非意圖描述樣本中個別入仕的具體情況，而是要獲得有關總體的統計資料。個別作答者因此決不會被識別出來，調查結果只會以集體概括形式發表，如統計圖、表等。有了這樣的保證，被訪者會較願意合作和提供真實答案。

在香港，普查及統計條例保障了政府統計處所蒐集資料的機密性。此外，不少從事或利用統計調查方法的專業組織都有一套道德守則，訂定確保調查所得資料機密性的規例。從事統計調查的組織，可參考下列一些保障調查資料機密性的辦法。

- (1)問卷上只用密碼識別被訪者。
- (2)拒絕向顧客或任何外間組織和人仕提供個別被訪者的具體資料。
- (3)當回收的資料已適當處理和存檔後，銷毀所有問卷。
- (4)供分析用的電腦儲存資料，不能有被訪者的姓名和地址。
- (5)所發佈的統計表，表內採用的分類要夠概括，以免關乎個別被訪者的資料能被技巧地推演出來。

本章所討論的內容，可參考 Moser and Kalton (Chapters 2, 7, 11, 12, 15) 及 Babbie (Chapters 9, 10, 12)。

第五章

資料分析

5.1 多樣的分析方法

資料校對完畢後便可進行分析工作。簡化及匯編資料可說是第一個重要步驟。隨之可進行調查分析的兩個主要部份，即「統計描述」(statistical description) 及「統計推論」(statistical inference)。適用於調查分析的統計方法可說多種多樣，於此實難一一詳述。測量變項間「關聯」(association) 的一般方法，及「估計」(estimation) 和「假設檢驗」(hypothesis testing) 等基本原理，均載於一般統計學教科書內（如 Freund, 1988），此處從略。

抽樣統計調查的主要目的，通常是估計若干總體中的「參數值」(parameters)；但不同的樣本設計要應用不同的計算公式。下文將簡述若干基本公式以供參考。

5.2 基於簡單隨機樣本的估計

通常最需要的總體參數值是平均值 \bar{X} ，總值 X_T ，及比例 P 。從樣本得出的參數值估計，很難絕對準確。所以一個估計的精確程度如何，應附有若干說明。常見的做法是附上一個估計的標準誤差。但標準誤差只能反映的情況是，在一有 N 個單位的總體，抽選 n 個單位作為樣本時，因隨機抽樣變異而導致的估計誤差。其他的誤差來源，應還包括無回應，以及在量度、校對、編碼和表列時的誤差（可參考 Coehran, chapter 13）。

若用簡單隨機抽樣方式，總體平均值 \bar{X} 的最常用估計是樣本平均值 \bar{x} ，式為

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

其中 x_1, \dots, x_n 均為樣本觀察值， \bar{x} 的標準誤差為

$$\text{s.e. } (\bar{x}) = \sqrt{(1 - f)} \frac{S}{\sqrt{n}}$$

其中 $f = n/N$ ，即「抽樣比」(sampling fraction)，而 S^2 則為「有限總體」(finite population) 的「方差」(variance)，式為

$$S^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}. \quad (5.1)$$

此外，因子 $1-f = (N-n)/N$ 稱為「有限總體修正因子」(finite population correction)。在應用上，當抽樣比 f 不超過 5% (有時甚至高達 10%) 時，有限總體修正因子可以略去。而省略此因子會高估 \bar{x} 的標準誤差。當應用「回置方式抽樣」(sampling with replacement) 時，此因子並不需要，而 f 則視作零。

計算標準誤差時涉及總體方差 S^2 ，但實際上這是一個未知數，只能用下式估計，

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}. \quad (5.2)$$

因而 \bar{x} 的標準誤差估計為：

$$s(\bar{x}) = \sqrt{1-f} \frac{s}{\sqrt{n}}.$$

這樣的估計有輕微偏差，但在大部份的應用範圍內，這偏差不算嚴重。

通常可以假設 \bar{x} 服從「正態分佈」(normal distribution)，而分佈的平均值為 \bar{X} 。 \bar{X} 的置信下限和上限則分別為：

$$\bar{x} - z \times s(\bar{x}) \quad \text{和} \quad \bar{x} + z \times s(\bar{x}),$$

式中 z 乃與所要求的置信水平 (confidence level) 相應的正態變量，最常採用的值為：

置信水平 (%)	50	80	90	95	99
z	0.67	1.28	1.64	1.96	2.58

如樣本大小 n 少於 50，則上述的百分點可取自 t 表，以 $(n-1)$ 為「自由度」(degree of freedom)。 t 分佈只在觀察值 x_i 成正態分佈，而 N 為無限大時才能確立。但稍偏離正態並不會有太大影響。若總體偏態程度大而樣本又小時，便要用特別方法處理了。

計算 \bar{x} 的標準誤差有三個目的：(1)用以比較由簡單隨機抽樣與其他抽樣法所得估計的精確度，(2)依 2.3 節內討論的方法去估計所需樣本大小，(3)估計已完成的調查所能達到的精確度。

有關上述的第二個目的，需要用到總體方差 S^2 ，但關乎這方差的估計，即樣本觀察值的方差 s^2 在調查之前尚屬未知數，故要用其他方法取得估值。如果有人曾做過類似總體及類似特質的調查研究，則此等調查所得的方差量度可作為 S^2 的參考。如在正式抽樣調查前曾進行試查，則其結果亦可用作 S^2 值的指引。一個較可靠的方法是先取一個初步簡單隨機樣本，其大小為 n_1 ，以估計 S^2 ，然後決定樣本大小。並以簡單隨機方式補足所欠缺的數目，即 $(n - n_1)$ 。通常有關總體結構的資料都可以幫助猜度 S^2 值。例如調查所關注的變項屬泊松分佈 (Poisson distribution)，則可用任何 \bar{x} 的暫估數值去估計 S^2 。

常用以估計總體中總值 X_T 的簡單隨機樣本估計是

$$x_T = N\bar{x}$$

其標準誤差

$$\text{s.e. } (x_T) = N\sqrt{1-f} \frac{S}{\sqrt{n}}$$

可用下式估計

$$s(x_T) = N\sqrt{1-f} \frac{s}{\sqrt{n}}$$

式 (5.1) 及 (5.2) 為 S 及 s 的定義。理論上，當 f 小時，用於估計 \bar{X} 的有限總體修正因子 $(1-f)$ 可以略去。而 X_T 的置信下限和上限分別為

$$x_T - z \times s(x_T) \quad \text{和} \quad x_T + z \times s(x_T)$$

\bar{x} 及 x_T 為 \bar{X} 及 X_T 的「無偏差估計」(unbiased estimate)。有關其特性及詳細理論可參閱 Cochran, Chapter 2。

總體比例 P 可用樣本比例 p 作估計， p 的標準誤差為

$$\text{s.e. } (p) = \sqrt{1-f} \frac{S}{\sqrt{n}}$$

此處 S 由下式決定

$$S^2 = \frac{N P(1 - P)}{N - 1}.$$

因 P 未知，所以 S^2 亦屬未知。 S^2 的無偏差估計為

$$s^2 = \frac{n p(1 - p)}{n - 1}$$

p 的標準誤差因此為

$$\begin{aligned}s(p) &= \sqrt{1 - f} \frac{s}{\sqrt{n}} \\&= \sqrt{1 - f} \sqrt{\frac{p(1 - p)}{n - 1}}.\end{aligned}$$

若 f 值小，有限總體修正因子可省略，故上式可簡化為

$$s(p) = \sqrt{\frac{p(1 - p)}{n - 1}}.$$

當 n 值不太小時，可利用正態分佈算出 P 的置信下限和上限的近似值為

$$p - z \times s(p) - \frac{1}{2n} \quad \text{和} \quad p + z \times s(p) + \frac{1}{2n}. \quad (5.3)$$

式中 $\frac{1}{2n}$ 是「連續校正數」(correction for continuity)，一般未經校正的正態近似值會提供較狹窄的置信間距。有關樣本比例的細節可參看 Cochran, Chapter 3。

5.3 基於較複雜樣本設計的估計

對於抽樣設計上比較複雜的樣本，上節提供的公式要經過修改才能適用。適用於簡單分層抽樣，簡單整羣抽樣和等距抽樣的公式，現簡述如下。有關細節及其他樣本設計方面，讀者可參閱 Cochran 一書。

下列的所有大寫字母均代表總體參數值，而小寫字母則代表樣本統計值。

「簡單分層抽樣法」(simple stratified random sampling)——即於總體中每層以簡單隨機抽樣法抽取樣本。

銓釋

L	總體所含層的數目
N_h	第 h 層中單位的數目 故 $N_1 + \dots + N_L = N =$ 總體中單位的數目
$W_h = \frac{N_h}{N}$	h 層的權數 (層與總體的相對大小)
n_h	第 h 層的樣本大小 故 $n_1 + \dots + n_L = n =$ 樣本大小
X_{hi}	第 h 層的第 i 個單位的數值
x_{hi}	第 h 層的第 i 個被選出單位的數值
$\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$	第 h 層的總體平均值
$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$	第 h 層的樣本平均值
$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2$	第 h 層的方差
$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$	第 h 層的樣本方差
P_h	第 h 層的總體比例
p_h	第 h 層的樣本比例

以簡單分層抽樣所得的 \bar{x}_{st} 估計 \bar{X} ，則式為

$$\bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h$$

$$\text{s.e. } (\bar{x}_{\text{st}}) = \sqrt{\sum_{h=1}^L W_h^2 (1-f_h) \frac{s_h^2}{n_h}}$$

其中

$$f_h = \frac{n_h}{N_h}$$

$$s(\bar{x}_{\text{st}}) = \sqrt{\sum_{h=1}^L W_h^2 (1-f_h) \frac{s_h^2}{n_h}}.$$

以 x_T 估計 X_T ，式為

$$x_T = N \bar{x}_{\text{st}}$$

$$\text{s.e. } (N \bar{x}_{\text{st}}) = N \times \text{s.e. } (\bar{x}_{\text{st}})$$

$$s(N \bar{x}_{\text{st}}) = N \times s(\bar{x}_{\text{st}}).$$

以 p_{st} 估計 P ，式為

$$p_{\text{st}} = \sum_{h=1}^L W_h p_h$$

$$\text{s.e. } (p_{\text{st}}) = \sqrt{\sum_{h=1}^L W_h^2 \frac{N_h - n_h}{N_h - 1} \frac{P_h(1-P_h)}{n_h}}$$

$$s(p_{\text{st}}) = \sqrt{\sum_{h=1}^L W_h^2 \frac{N_h - n_h}{N_h - 1} \frac{p_h(1-p_h)}{n_h}}.$$

「簡單整羣抽樣法」 (simple cluster sampling) —— 在有 M 羣的總體中，以簡單隨機抽樣法抽出 m 羣，所有被抽選中的羣內單位都包括在樣本內。

銓釋：

M 總體所含羣的數目

m 樣本所含羣的數目

N_i 第 i 個羣中單位的數目

故 $N_1 + \dots + N_m = N$ = 總體中單位數目

n_i 第 i 個被抽選中的羣中單位的數目

故 $n_1 + \dots + n_m = n$ = 樣本大小

X_{ij} 第 i 個羣的第 j 個單位的數值

x_{ij} 第 i 個被抽選中的羣的第 j 個單位的數值

X_i 第 i 個羣的平均值

x_i 第 i 個被抽選中的羣的平均值

X_{iT} 第 i 個羣的總值

x_{iT} 第 i 個被抽選中的羣的總值

P_i 第 i 個羣的比例

p_i 第 i 個被抽選中的羣的比例

常用以估計 \bar{X} 的方法共有三個：

(a) 「整羣樣本比率」 (cluster sample ratio) \bar{x}_{ca}

$$\bar{x}_{ca} = \frac{\sum_{i=1}^m x_{iT}}{\sum_{i=1}^m n_i}$$

$$\text{s.e. } (\bar{x}_{ca}) = \sqrt{\frac{(M-m)M}{(M-1)m} \sum_{i=1}^m \left(\frac{N_i}{N} \right)^2 (\bar{X}_i - \bar{X})^2}$$

$$s(\bar{x}_{ca}) = \sqrt{\frac{(M-m)M}{(m-1)m} \sum_{i=1}^m \left(\frac{N_i}{N} \right)^2 (\bar{x}_i - \bar{x}_{ca})^2}.$$

如 N 為未知數，則可用 Mn/m 估值代替。

(b) 「整羣樣本總值」 (cluster sample total) \bar{x}_{cb}

$$\bar{x}_{cb} = \frac{M}{Nm} \sum_{i=1}^m x_{iT}$$

$$s.e. (\bar{x}_{cb}) = \sqrt{\frac{(M-m)M}{(M-1)mN^2} \sum_{i=1}^M (X_{iT} - \bar{X}_T)^2}$$

$$s(\bar{x}_{cb}) = \sqrt{\frac{(M-m)M}{(m-1)mN^2} \sum_{i=1}^m (x_{iT} - \bar{x}_{cb})^2}.$$

(c) 「樣本羣平均值的不加權平均值」 (unweighted average of the chosen cluster means) \bar{x}_{cc}

$$\bar{x}_{cc} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

$$s.e. (\bar{x}_{cc}) = \sqrt{\frac{M-m}{mM(M-1)} \sum_{i=1}^M (\bar{X}_i - \bar{X}_c)^2}$$

其中 $\bar{X}_c = \frac{1}{M} \sum_{i=1}^M \bar{X}_i$

$$s(\bar{x}_{cc}) = \sqrt{\frac{M-m}{mM(m-1)} \sum_{i=1}^m (\bar{x}_i - \bar{x}_{cc})^2}.$$

上述三個估計，即 \bar{x}_{ca} ， \bar{x}_{cb} ， \bar{x}_{cc} 的比較如下：

估計量		
\bar{x}_{ca}	\bar{x}_{cb}	\bar{x}_{cc}
有無偏差	有偏差；只在 m 很大時偏差	無偏差。 N_i 之間差別不大時，偏差不嚴重，除非羣平均值與羣的大小有強相關。
標準誤差	視羣平均值間差別而定，差別小則標準誤差亦較小。	常較 \bar{x}_{ca} 大；除非 N_i 之間差別不大。計算預期均方 error) 比標準誤差較適當；但如 N_i 之間差別小則當別論。
如何估計 X_T	$N\bar{x}_{ca}$	$N\bar{x}_{cb}$ $= \frac{M}{m} \sum_{i=1}^m x_{iT}$
相對優點	較高效率	無偏差，估計 X_T 時無需知道 N 簡單、快捷

以 p_c 估計 P ，式為

$$p_c = \frac{M}{Nm} \sum_{i=1}^m N_i p_i$$

$$s.e. (p_c) = \sqrt{\frac{M-m}{mM(M-1)\bar{N}^2} \sum_{i=1}^M (N_i P_i - \bar{N} P)^2}$$

其中 $\bar{N} = \frac{1}{M}(N_1 + \dots + N_M)$

$$s(p_c) = \sqrt{\frac{M-m}{mM(m-1)} \sum_{i=1}^m \left(\frac{M}{N} N_i p_i - p_c \right)^2}.$$

「等距抽樣法」 (systematic sampling) ——亦稱系統抽樣法，是以等間距方式自一完整總體清單中抽取樣本單位。

一個等距樣本可被視為一個總體中有 M 個羣的整羣樣本，各羣的單位分別為：

$$\begin{aligned} \text{第一羣 } & X_1, X_{M+1}, X_{2M+1}, \dots \\ \text{第二羣 } & X_2, X_{M+2}, X_{2M+2}, \dots \\ & \vdots \\ \text{第 } M \text{ 羣 } & X_M, X_{2M}, X_{3M}, \dots \end{aligned}$$

每羣的單位數目為 n 或 $n + 1$ ，而樣本中只有一個羣 ($m = 1$)，這個羣是在 M 個羣中以隨機方法抽出。但如總體清單的排列屬隨機性，則等距樣本可作為簡單隨機樣本處理。其他細節可參閱 (Cochran, Chapter 8)。

上文討論的，僅是各種較簡單的抽樣設計中的基本公式。在實際工作中，可能遇到較複雜的情況，涉及「多段抽樣」 (multi-stage sampling)、「伴隨變項」 (concomitant variable)、成本因素等個別或綜合問題。有關的技術細節，讀者可參看如 Cochran 等書籍或向專業人士請教。

5.4 注意分組後的樣本大小

統計調查常有多重的目的，要同時研究相當數目的變項。估計總體值時不但會關乎全總體，同時更會牽涉到一系列的組別；如不同年齡組或不同教育水平的人口，不同的地理區域等等。

一個本來可以提供可靠估計的樣本，當劃分得過份精細時，往往無法用來進行分組分析。因為過份精細的分組，其「實際樣本大小」 (effective sample size) 會相當小，變項的抽樣誤差便會較大。故分析調查結果時，要小心選定適當的分組方法和準則。

較大的樣本可容許進行較精細的分組分析，但當決定樣本大小時，應同時考慮當樣本增大時，所增加的成本相對於所提高的分析效益是否值得？

5.5 足夠高的回應率的重要性

就某一個調查總體而言，調查回應率是指合適基本單元所完成的問卷數與樣本中合適基本單元總數之間的比率。按此定義，則所有空白和架構外單元等（見 §2.5）非合適基本單元均不計算在分子或分母之內。

低回應率對調查結果的不利影響，較之一個細小樣本更嚴重。這是因為對調查中的無回應者，並無推斷其特質的有效科學方法。故此如要對研究中的總體作出有效估計，便先要有一個足夠高的回應率。低回應率的調查，其結果會有誤導性。

不能取得樣本中基本單元的資料（即無回應）是統計調查的一個嚴重問題。因為無回應者與作答者可能在調查中的變項上存在差別，所以用低回應率的調查資料去作估計時，便難免有所偏差。這種「無回應偏差」 (non-response bias) 可說明如下。

設若某情況下總體可分為兩個部份，即作答者和無回應者。假如調查的目的是要決定總體平均值 M ，式為

$$M = W(r)M(r) + W(m)M(m)$$

其中 $M(r)$ 及 $M(m)$ 分別為作答者及無回應者的組平均值（ r 為作答者， m 為無回應者）， $W(r)$ 和 $W(m)$ 分別代表兩組單元在總體中所佔比例，〔即 $W(r) + W(m) = 1$ 〕。因調查中無法取得無回應者資料，故有關的估計只能是 $M(r)$ ，而 $M(r)$ 與總體平均值 M 的差為

$$M(r) - M = W(m) [M(r) - M(m)]$$

這個差別便是無回應偏差，是由於採用作答者的組平均值代替總平均值而產生。差別的大小視乎兩個因子： $W(m)$ ，即無回應者佔總體中比例；及 $[M(r) - M(m)]$ ，亦即作答者及無回應者兩個組平均值的差。

假如作答者和無回應者兩組的平均值是相等或十分接近，則無回應偏差便不存在或只是十分輕微。但實際上，把缺失的回應視作隨機的缺失是不適當的假定。而無回應者的特質往往與作答者不同。故此唯一能確實減低無回應偏差的辦法是盡量減少無回應者的數目，即要求小的 $W(m)$ 。

將樣本加權以至配合已知的總體分佈有助調整無回應及處理涵括不全的問題，後者指調查總體中的若干基本單元未被包括在抽樣架構內。舉例說，最近的人口普查資料可提供總體的年齡分佈。那麼在一項調查中，若年青人口的無回應率偏高，或抽樣架構遺漏了不少年青人口，則樣本可透過加權方法，使之符合已知的年齡分佈，以補償上述的缺失。但如每個年齡組內，作答者及無回應者間仍在調查變項上有一定的差距，則若干無回應偏差仍難消除。

作答者的資料，亦可能有缺漏。部份被訪者對若干問題可能沒法作答，或因問題過於敏感或尷尬，又或認為問題不切合調查目標而拒絕作答；訪問員亦可能錯誤地跳問問題或漏寫答案而造成缺漏。即使問題的答案已記錄在問卷上，亦會因與其他的答案矛盾而在校對時被剔除。這些被訪者資料紀錄中的缺漏稱為「項目無回應」(item non-response)。

項目無回應率的定義是：對某一問題沒有作答的合適基本單元數目除以樣本中合適基本單元的總數。與無回應的加權調整方法相比，各種設算的方法，都是嘗試對項目無回應作出補償。設算的一個主要好處是建立一個無遺漏數值的被訪者資料集，有利分析調查結果。

如真遇到極低回應率的情況，統計調查的負責人應準備接受這個事實並承認調查失敗，更不應發表調查的結果。不顧回應率便隨便公佈調查結果是不負責任的行為。即使在報告中知會讀者亦未算足夠。

本章討論的內容，可參考 Babbie (Chapter 16)。

第六章

報導統計調查結果

一份完整的統計調查報告應詳細報導有關調查各方面的細節；特別是調查總體、樣本設計、抽樣誤差、回應率及非抽樣誤差的可能成因等的具體情況。此外更應附有原來問卷的樣版。技術細節亦應一起報導，可選擇放在附錄內。研究人員宜就讀者類別，選擇適當語言、體裁及表達方式來撰寫調查報告。

當向大眾傳播媒介發佈涉及公眾利益或衆所關注事項的調查結果時，負責人員在報導調查發現之外，更應充份提供調查方法的詳細資料，以便傳播媒介能一起報導這兩方面的資料，否則社會大眾便無從判別調查結果的可靠性。

調查設計或許有偏差，事實上這種情況有時是無法避免的，那麼調查人員應向公眾指出設計上的缺點，和所導致的後果。一般會出現的設計偏差包括有樣本選擇偏差，加權偏差，問題偏差和問題編排次序所造成的影响等。通常調查人員會知悉這些問題所在。當有這些偏差存在時，調查所得結論的局限亦要清楚說明。

舉例說，在不得已情況下採用非概率樣本進行調查時，應清楚指出抽樣法上的可能偏差，並小心避免不適當地作出有關較大總體的推論。

不少統計調查報告或關於統計調查的報導，都傾向於略去有關方法和技術的部份。作者或許認為讀者對這部份不感興趣或無法了解。雖然事實可能的確如此，但省略了這些卻極可能誤導讀者。故此作者無論如何也應以簡單、易明、不枯燥的方式說明所用方法和技術，一般大眾亦會因此而加深對統計及常用研究法的認識。

當報導民意調查結果時，不應忽略下列各項資料：

- (1) 統計調查的主辦機構；
- (2) 資料回收的期間；
- (3) 抽樣方法；
- (4) 蒐集資料方式（見 §4.1）；
- (5) 問題的措詞；
- (6) 總體的涵括範圍；
- (7) 樣本大小及回應率；
- (8) 作「分組分析」(subgroup analysis) 時，有關樣本的大小及回應率；
- (9) 盡可能說明誤差大小或「置信區間」(confidence interval)；
- (10) 非抽樣誤差的可能成因。

上述討論可參閱 Babbie (Chapters 18, 19, 20) 及 Hage et al (104—105 頁)。

參考文獻

- Babbie, E. (1990). *Survey Research Methods*. 2nd ed. Wadsworth.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley and Son.
- Freund, J.E. (1988). *Modern Elementary Statistics*. 7th ed. Prentice-Hall.
- Hage, G., Dennis, E., Ismach, A. & Hartgen, S. (1976). *New Strategies for Public Affairs Reporting*. Prentice Hall.
- Kalton G. (1983). *Introduction to Survey Sampling*. Sage Publication.
- Moser C.A. & Kalton G. (1971). *Survey Methods in Social Investigation*. Heinemann Educational Book Ltd. E.L.B.S. edition of 2nd ed.
- Smith, H.W. (1975). *Strategies of Social Research — the Methodological Imagination*. Prentice Hall.