

抽 樣

Chapter 6

Introduction to Social Research 5E

研究方法

二版

基礎理論與技巧

Earl Babbie 原著 林佳瑩 審閱
蔡毓智、邱泯科、陳佳穎、姜馨彥 合譯



6.1 引言

- 調查抽樣最常見的用途之一是用在政治或選舉的民意調查 (political polling)。
- 社會研究者必須選取觀察對象，藉以概推其他無法被觀測之人群與事件。
 - 選取觀察對象的過程，就稱為 抽樣(sampling)。
- 了解抽樣的邏輯對進行社會研究是必要的。



6.1 引言

表 6-1 2008 年投票前夕各媒體的投票意向調查

調查單位	截止日	Obama	McCain
FOX	Nov 2	54	46
NBC/WSJ	Nov 2	54	46
Marist College	Nov 2	55	45
Harris Interactive	Nov 3	54	46
Reuters/C-SPAN/Zogby	Nov 3	56	44
ARG	Nov 3	54	46
Rasmussen	Nov 3	53	47
IBD/TIPP	Nov 3	54	46
DailyKos.com/Research 2000	Nov 3	53	47
GWU	Nov 3	53	47
Marist College	Nov 3	55	45
審查結果	Nov 4	54	46



6.1 引言

本圖顯示了各個獨立調查的資料都產生了事實的相同結果。它也顯示了國家危機對總統的支持度的衝擊：在本例中為 911 事件發生前後布希總統的支持度變化。

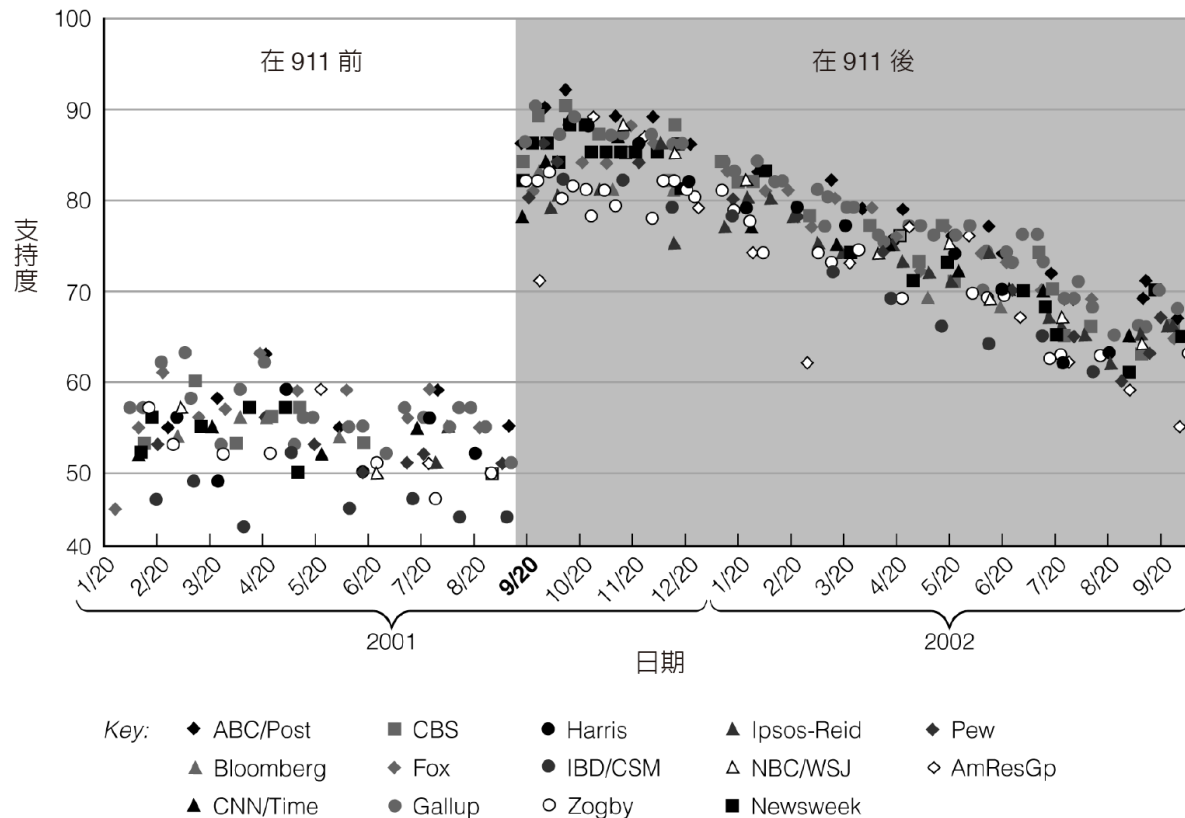


圖 6-1 布希總統的支持度：原始民調資料

本圖顯示了各個獨立進行的民調結果產生了相同的事實圖像。它也顯示了全國性的危機對總統支持度的影響。本例中是 911 恐怖攻擊前後總統支持度的變化。

6.2 抽樣的簡史

- 社會研究領域中，抽樣技術的發展是與選舉民意調查攜手並進的。
- 6.2.1 Alf Landon 總統
 - Digest 的民意調查—抽樣架構 (sampling frame) 可能有問題。
- 6.2.2 Thomas E. Dewey 總統
 - Gallup 民意調查—定額抽樣法 (quota sampling)。
- 6.2.3 兩種類型的抽樣方法
 - 非機率抽樣、機率抽樣



6.3 非機率抽樣

- 非機率抽樣 (nonprobability sampling)
 - 任何使用不是經由機率理論所主張的方式所選取樣本的技術，均可稱為非機率抽樣。
- 6.3.1 仰賴可得受訪者法 (reliance on available subjects)
 - 常被稱為「便利性」或「偶然」抽樣。
 - 對社會研究而言，此方法對於樣本的代表性，無法提供任何的 control。是極端危險的抽樣方法。



6.3 非機率抽樣

□ 6.3.2 立意或判斷抽樣法 (purposive or judgmental sampling)

□ 立意抽樣基於研究者判斷何者為最適當或最具代表性，選取所觀察的單位作為樣本。

□ 6.3.3 滾雪球抽樣法 (snowball sampling)

□ 通常在田野研究中使用，當訪問某一個受訪者時同時請他建議其他受訪的對象。有時被認為是一種偶遇抽樣(accidental sampling)。



6.3 非機率抽樣

□ 6.3.4 定額抽樣法 (quota sampling)

- 定額抽樣基於預設的特徵基礎上進行樣本的選取，因此全部的樣本會與被研究的母體中預設存在的特徵具備相同的特徵分配。
- 以矩陣(matrix)、或表格的方式，陳述目標母體的特徵。



6.3 非機率抽樣

□ 6.3.5 選取「報導人」

- 當研究者進行一項實地研究，企圖了解某些社會情境 (social setting) 時，他們的了解和知識，大部分來自於被研究團體中某些成員的合作。
- 報導人 (informant)：
 - 某些人對你所想要研究的社會現象有相當的了解，同時也願意告訴你他所知道的現象。
 - 請不要將報導人與受試者二者產生混淆。



6.4 機率抽樣的理論與邏輯

□ 機率抽樣 (probability sampling)

- 依據機率理論而進行樣本選取的一般型式，典型的機率抽樣涉及了某些隨機選取的機制。特定的機率抽樣的型式包括了EPSEM, PPS, 簡單隨機抽樣及系統抽樣等方式。
- 基本意涵
 - 為了能夠對整個母體提供有用的描述，一個來自整個母體的樣本個體，必須包含存在於母體的同樣變異。



6.4 機率抽樣的理論與邏輯

6.4.1 有意與無意的抽樣偏誤

偏誤 (bias)：意謂所選取的樣本無法典型地代表其母體。

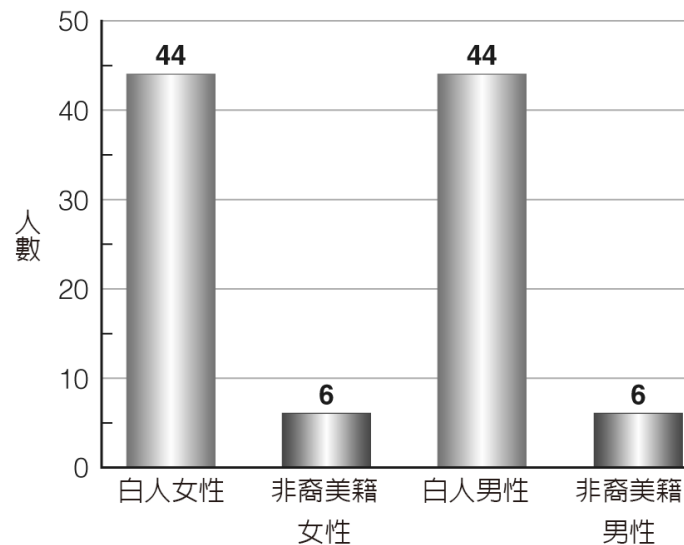


圖 6-2 100 個民衆的母體

正常的情况下，抽樣的目的是反映較大的母體的特徵與動態。本圖是為了展示的目的而描繪，本例中假設了母體成員只有 100 個人的情况。



6.4 機率抽樣的理論與邏輯

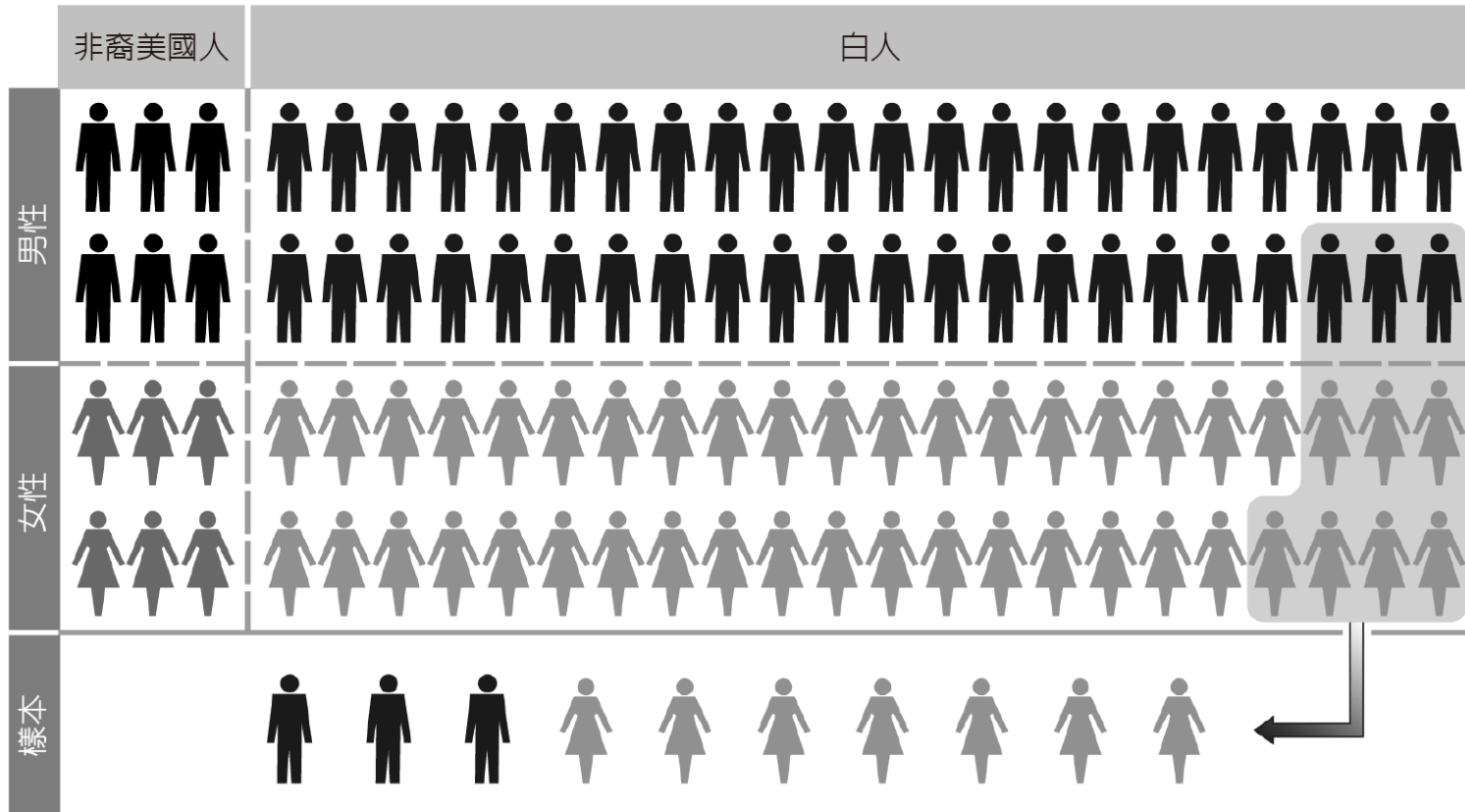


圖 6-3 一個便利性取樣的例子：方便，但是不具代表性。

選取並觀察手邊既有的人或許是最容易的方法，但此種方法較不可能產生能夠反映母體真實情況的樣本。



6.4 機率抽樣的理論與邏輯

6.4.2 代表性與抽取機率

代表性 (representativeness)

- 樣本與樣本被選取的母體具有相同特徵分配的品質。
- 代表性的意涵指由樣本所作出的描述與解釋可被預設為與具有母體相同的特徵。
- 代表性可藉由機率抽樣的方式增強，並且提供概推性與推論統計使用。

EPSEM (equal probability of selection method)



6.4 機率抽樣的理論與邏輯

- 機率抽樣係以隨機選取(**random selection**) 的程序為基礎。
- **元素 (element)**
 - 組成母體以及被選入樣本的單位。元素與分析單位有別，後者是用在資料分析中。
- **母體(population)**
 - 一個研究中由理論所界定元素的聚集。
- **研究母體(study population)**
 - 樣本被實際選取的元素之聚集。



6.4 機率抽樣的理論與邏輯

□ 6.4.3 隨機選取 (random selection)

- 一種抽樣方法，此方法中每一個元素都有相同的被選取的機會，並獨立於選取過程中的任何其他事件。
- 抽樣的最終目的：
 - 從母體抽取若干元素，而我們對這些元素的描述，可以正確的描述其所來自的母體。
- 抽樣單位 (sampling unit)
 - 在某些抽樣的階段中所考慮選取的元素或元素組(set)。



6.4 機率抽樣的理論與邏輯

- 6.4.4 機率理論、抽樣分配、及抽樣誤差的估計
 - 機率理論 (probability theory)
 - 數學的一個分支，提供了估計母體參數的基礎。
 - 抽樣分配 (sampling distributions)
 - 以一個從母體抽出的單獨樣本，我們可以估計母體的參數。
 - 抽樣誤差 (sampling error)
 - 在機率抽樣中誤差被期待的程度。
 - 決定抽樣誤差的公式三個因素：
 - 樣本規模、母體參數、以及標準誤。



6.4 機率抽樣的理論與邏輯

□ 十個個案的抽樣分配

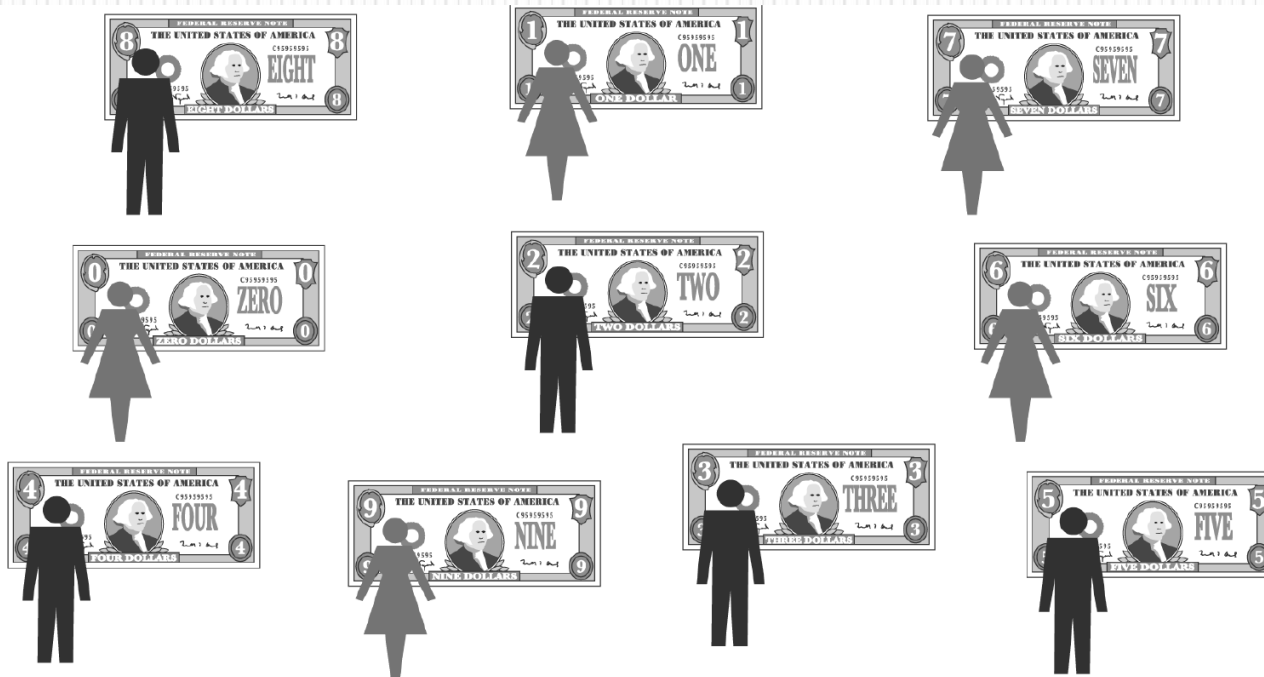
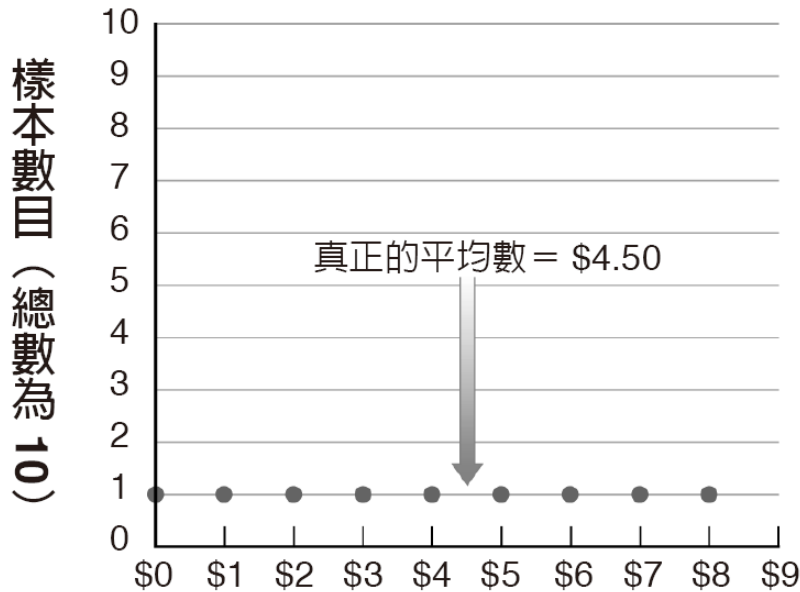


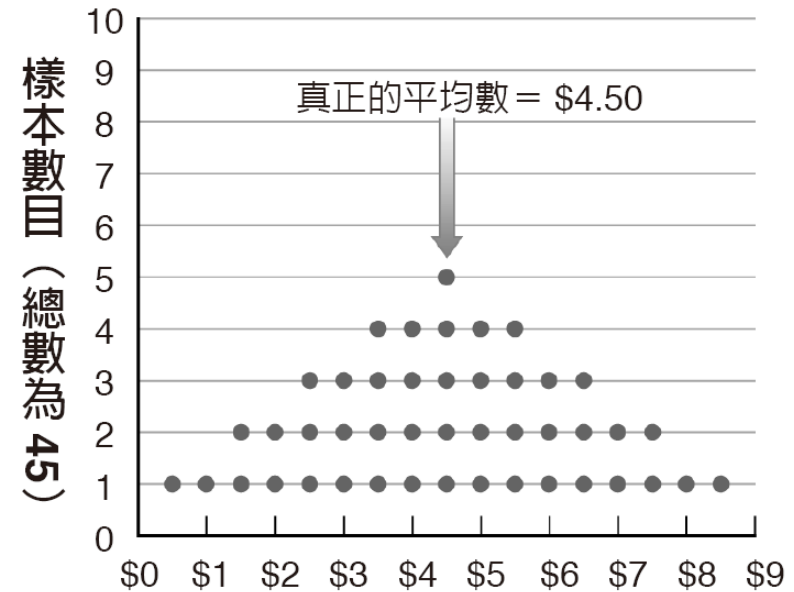
圖 6-4 擁有美金 0-9 元的 10 個人的母體。
讓我們假設一群只有 10 個人的母體，他們的口袋中擁有從 0 到 9 元不等的金錢。



6.4 機率抽樣的理論與邏輯



平均數的預測值（樣本規模為 1）



平均數的預測值（樣本規模為 2）

圖 6-5 樣本規模為 1 的抽樣分配。
在這個簡單的例子中，人們所擁有的金錢平均為 4.5 元，假如我們選取 10 個樣本數為 1 的樣本，我們對平均數的「估計」將會包含所有的範圍。

圖 6-6 樣本規模為 2 的抽樣分配。
在我們將樣本規模增加到 2 之後，這些所有可能的樣本提供了平均數較佳的估計，估計值開始朝 4.5 元這個真正的平均值開始靠近。



6.4 機率抽樣的理論與邏輯

□ 抽樣分配與抽樣誤差的估計

□ 統計值 (statistic)

□ 一個樣本中某個變項的總結性描述，通常用來估計母體參數。

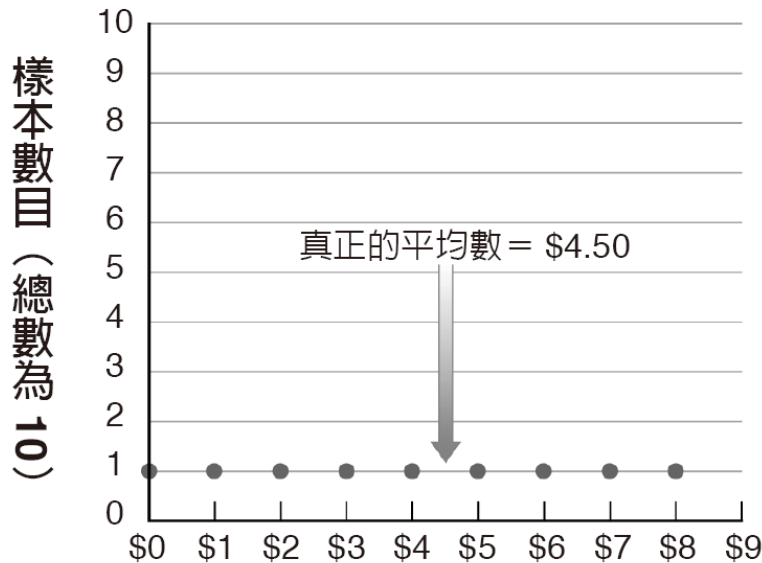
□ 標準誤 (standard error)：是抽樣誤差的測量值。

$$s = \sqrt{\frac{P \times Q}{n}}$$

□ 在機率理論裡，標準誤提供非常有價值的資訊。

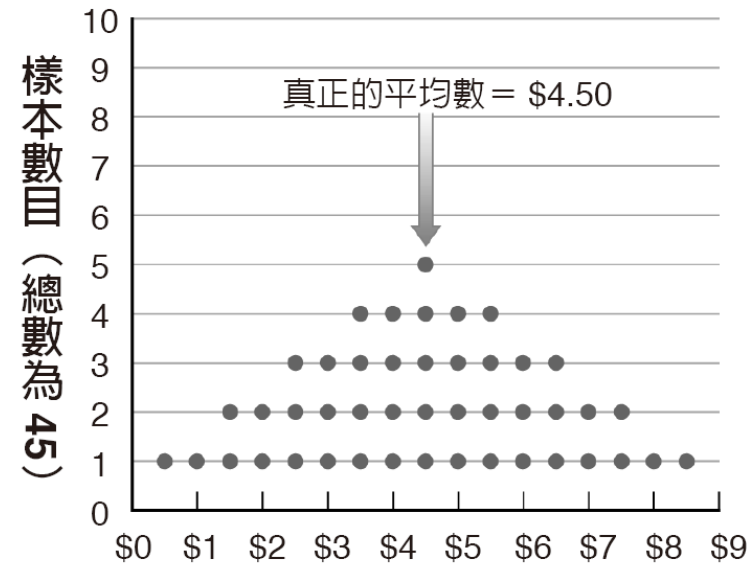


6.4 機率抽樣的理論與邏輯



平均數的預測值 (樣本規模為 1)

圖 6-5 樣本規模為 1 的抽樣分配。
在這個簡單的例子中，人們所擁有的金錢平均為 4.5 元，假如我們選取 10 個樣本數為 1 的樣本，我們對平均數的「估計」將會包含所有的範圍。



平均數的預測值 (樣本規模為 2)

圖 6-6 樣本規模為 2 的抽樣分配。
在我們將樣本規模增加到 2 之後，這些所有可能的樣本提供了平均數較佳的估計，估計值開始朝 4.5 元這個真正的平均值開始靠近。



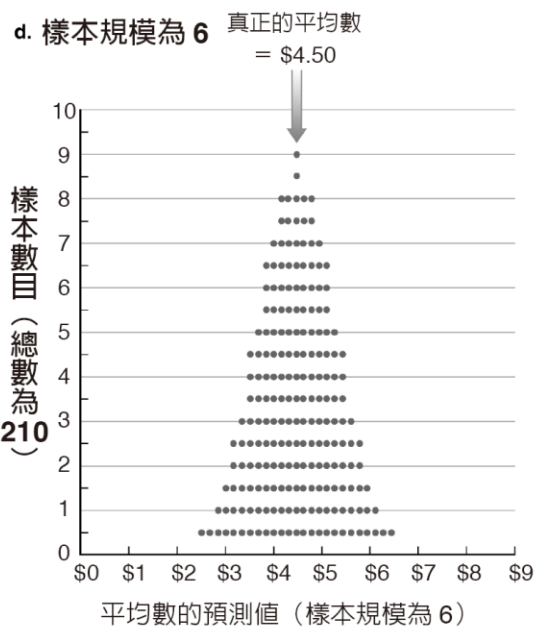
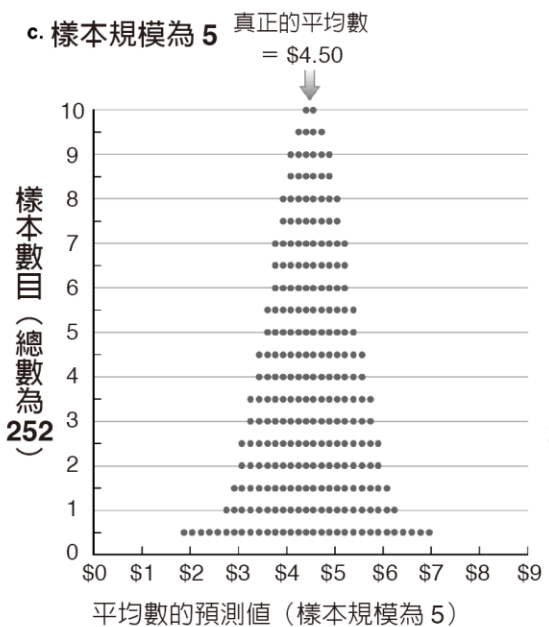
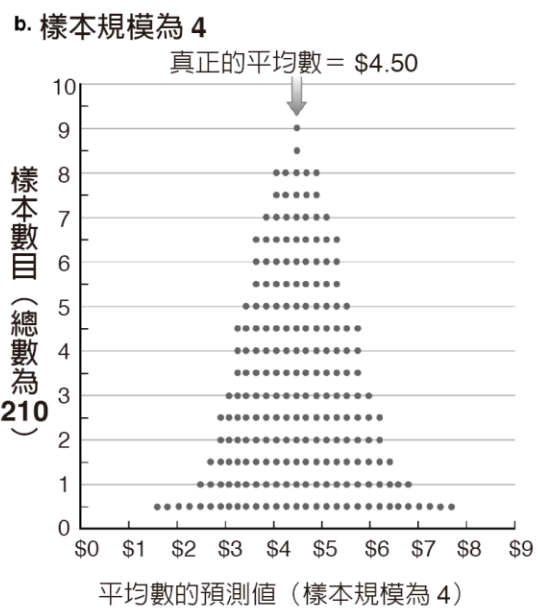
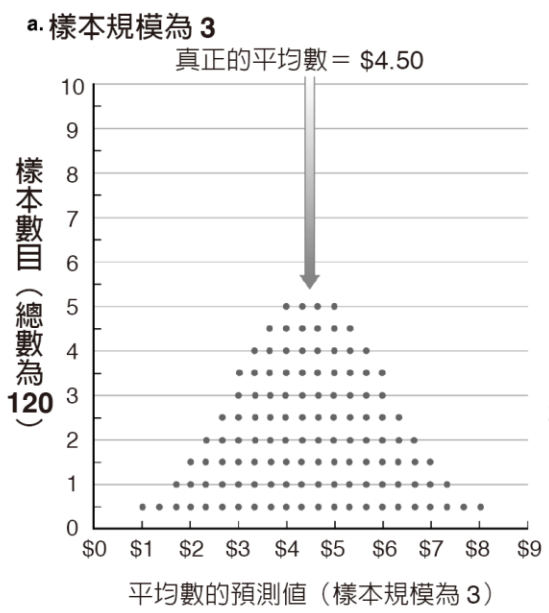


圖 6-7 樣本規模為 3、4、5、6 的抽樣分配。
 隨著我們加大樣本規模，可能的樣本平均值也愈來愈朝真正的平均值靠近。極端不可能估計的機會被減少到分配的兩端，接近真正數值的百分比也增加了。



6.4 機率抽樣的理論與邏輯



圖 6-8 可能樣本研究結果的範圍
轉換到更為真實的例子，讓我們假設我們想要了對關心學生守則的學生進行抽樣。讓我們假設有 50% 學生贊成，另外一半反對—雖然真正的情況研究者無從得知。

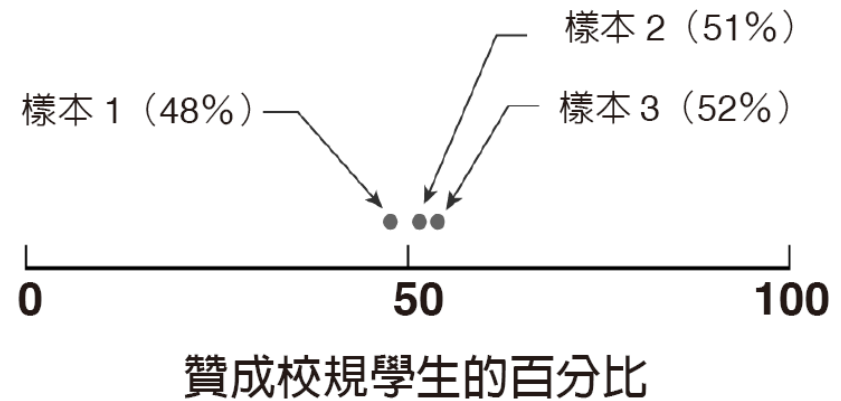


圖 6-9 三個假設性研究所產生的結果
假定有一個更大的學生群，假設我們選取了三個不同的樣本，每一個都具備可觀的樣本規模。我們並不必然指望這些樣本會正確反映整體學生群的態度，但它們應該會是相當接近的。



6.4 機率抽樣的理論與邏輯

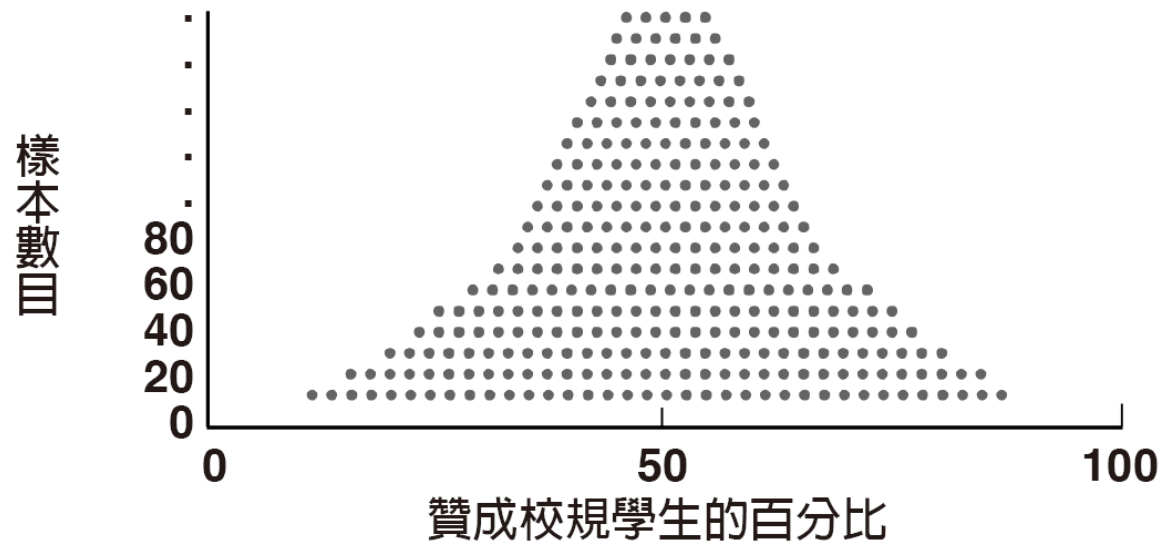


圖 6-10 抽樣分配

假如我們選取許多較大規模的良好樣本，我們會期望它們應該會相當接近真正的值 (50%)，在這些樣本中只有少數會落在標誌之外的地方。



6.4 機率抽樣的理論與邏輯

- 通常，我們不會知道母體參數是多少：
 - 我們之所以進行抽樣調查的理由，正就是為了估計該母體參數值。
- 而且我們也不會抽取很多的樣本：
 - 我們只會抽取一個樣本，我們可以針對實際抽取進行研究的那個單一樣本，做出若干預設。



6.4 機率抽樣的理論與邏輯

- 信心水準 (confidence level)
 - 母體參數會落入某個既定信賴區間的可估計機率。
- 信賴區間 (confidence interval)
 - 母體參數被估計可能落入的數值的範圍。
- 通常，會以下列方式表現樣本統計值的正確性：
 - 在某一信心水準之下，樣本統計值將會界於母體參數上下若干間距範圍之內。
- 以一個從母體抽取出來的樣本為基礎，我們可以估計母體的參數，以及估計可能的誤差程度。



6.4 機率抽樣的理論與邏輯

- 在少數的選取夠大比例的樣本的情況下，就可以計算一個「有限母體修正」以調整信賴區間。

$$\text{有限母體修正} = \sqrt{\frac{N-n}{N-1}}$$



6.5 母體與抽樣架構

□ 抽樣架構 (sampling frame)

- 構成母體單位的清冊或準清冊，從此清冊中得以選取樣本。
- 抽樣架構必須與我們希望研究的母體一致。
- 最簡單的抽樣設計裡，抽樣架構其實就是研究母體之所有元素的表列清冊。
- 實務上，通常研究者藉由既有的抽樣架構界定研究母體，而非由研究母體界定抽樣架構。

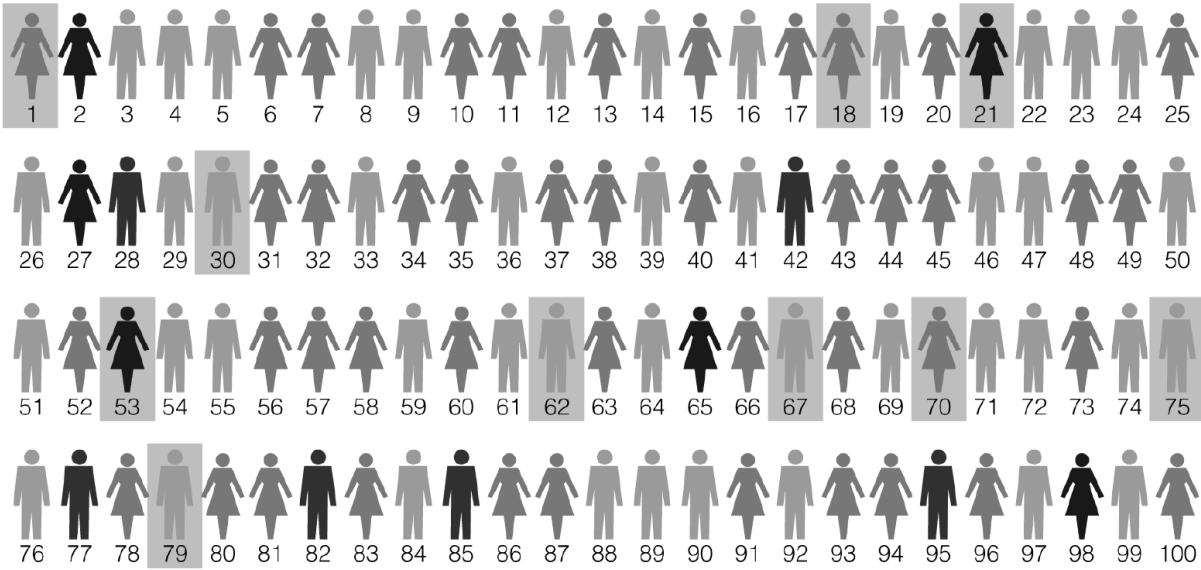


6.6 抽樣設計的類型

□ 6.6.1 簡單隨機抽樣 (simple random sampling, SRS)

- 一種機率抽樣的方式，在此種方式中，母體中的母一個單位均被賦與一個流水號，接著產生一組隨機號碼，具有這些號碼的單位就被選入樣本。





亂數表		
10480	15011	01536
22368	46573	25595
24130	48360	22527
42167	93093	06243
37570	39975	81837
77921	06907	11008
99562	72905	56420
96301	91977	05463
89579	14342	63661
85475	36857	53342
28918	69578	88231
63553	40961	48235
09429	93969	52636

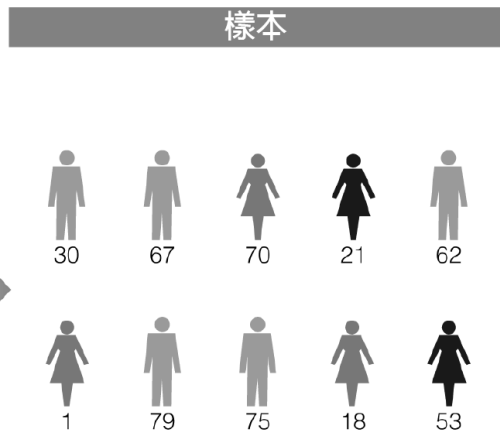


圖 6-11 一個簡單隨機樣本的例子
 假如我們將母體中的每一個人均加以編號，我們就可以使用隨機亂數表從整個母體中選取具代表性的樣本，任何具備從表中被選取的數字的人們就選為樣本。



6.6 抽樣設計的類型

□ 6.6.2 系統化抽樣 (systematic sampling)

- 一種機率抽樣的方式，在此種方法中，清冊中每隔 k 個單位就被選入到樣本中。
- 你可以將母體規模除上所要的樣本規模就會得出 k 值； k 被稱為取樣間隔。
- 在某些限制下，系統化抽樣與簡單隨機抽樣在功能上是相等的，並且較容易使用。通常第一個單位是由隨機方式選取的。



6.6 抽樣設計的類型

- 系統化抽樣中，經常使用兩個專有名詞：
 - 抽樣間距 (sampling interval)：
 - 母體中被選取為樣本的二個元素之間的標準距離。

$$\text{抽樣間距} = \frac{\text{population size}}{\text{sample size}}$$

- 抽樣比率 (sampling ratio)：
 - 母體中被選取為樣本的元素之比率。

$$\text{抽樣比率} = \frac{\text{sample size}}{\text{population size}}$$



6.6 抽樣設計的類型

□ 6.6.3 分層抽樣 (stratification)

- 構成母體的單位之集群，在抽樣前每個集群均為同質性的（或稱為層）。
- 這個程序可以結合簡單隨機、系統抽樣及集群 (Cluster) 抽樣來使用，有助於增進樣本代表性，至少就用以分層的變項而言是如此。
- 如何選取變項作為分層依據，通常是取決於手上可得的變項而定。



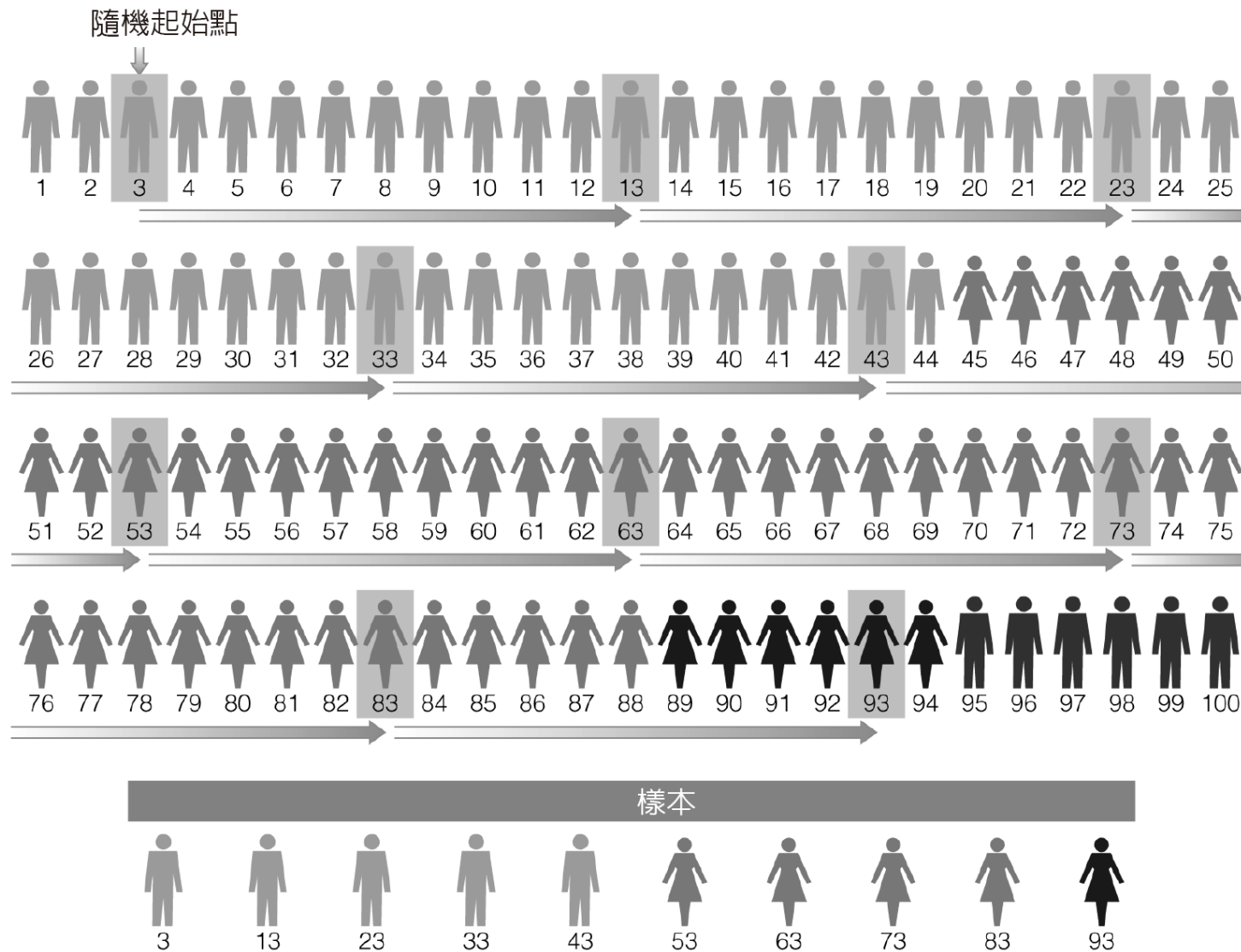


圖 6-12 一個分層、系統，以隨機起始產生的樣本。

一個分層的、系統化樣本涉及二個階段。首先，母體成員被分派到同質性的層級中；本例中僅僅以性別作為分層變項，你也可以使用更多的變項。接著在分層中間隔每 k 個人選出樣本。



6.6 抽樣設計的類型

□ 6.6.4 系統化抽樣所隱含的分層化

- 若母體清冊的排列方式，隱含存在分層化的模式，則系統化抽樣較簡單隨機更為正確。

□ 6.6.5 案例：大學生的抽樣

- 研究母體與抽樣架構
- 分層化
- 樣本選取
- 樣本的修改



6.7 多階段集群抽樣

□ 集群抽樣 (cluster sampling)

- 一種多階段抽樣，這種方式在一開始時選取自然的群體（集群），在隨後的抽樣中，這些被選取的群體的成員又被分為次樣本 (subsampled)。
- 多階段集群抽樣法其實是重複兩個步驟：
 - 表列清冊
 - 抽出樣本

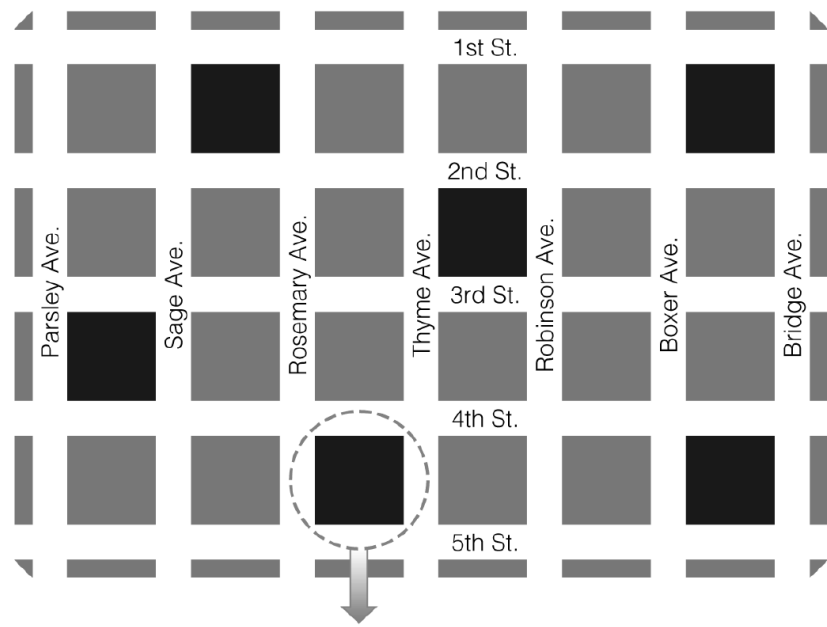


6.7 多階段集群抽樣

□ 6.7.1 多階段設計與抽樣誤差

- 多階段集群抽樣法深具高度效率，在此效率的代價，卻是獲得一個較不正確的樣本。
- 抽樣誤差可以藉由兩種因素加以降低：
 - 增加樣本規模
 - 增加所抽取元素之間的同質性





階段一：分辨街廓並選擇樣本（陰影部分代表被選出的街廓）

階段二：針對被選出的街廓，依序列出所有的家戶

1. 491 Rosemary Ave.
2. 487 Rosemary Ave.
3. 473 Rosemary Ave.
4. 455 Rosemary Ave.
5. 437 Rosemary Ave.
6. 423 Rosemary Ave.
7. 411 Rosemary Ave.
8. 403 Rosemary Ave.
9. 1101 4th St.
10. 1123 4th St.
11. 1137 4th St.
12. 1157 4th St.
13. 1169 4th St.
14. 1187 4th St.
15. 402 Thyme Ave.

階段三：在列出的每一街廓家戶中，抽取家戶的樣本（在這個例子中，隨機選取號碼5為起始點，每6個家戶印選取一戶作為樣本）

16. 408 Thyme Ave.
17. 424 Thyme Ave.
18. 446 Thyme Ave.
19. 458 Thyme Ave.
20. 480 Thyme Ave.
21. 498 Thyme Ave.
22. 1186 5th St.
23. 1174 5th St.
24. 1160 5th St.
25. 1140 5th St.
26. 1122 5th St.
27. 1118 5th St.
28. 1116 5th St.
29. 1104 5th St.
30. 1102 5th St.

圖 6-13 多階段集群抽樣。
 在多階段集群抽樣中，我們從選取集群樣本開始（在本例中為街廓）。接著我們將集群中的人任製作清冊（本例中為家戶），然後從被選取的集群中選出元素樣本。



6.7 多階段集群抽樣

□ 6.7.2 多階段集群抽樣中的分層

- 分層化可以應用在每一階段的抽樣中。其主要目的，即是達到更高的同質性。

□ 6.7.3 按規模之比例的機率抽樣(probability proportionate to size，簡稱PPS)

- 此種方法中，被選取的集群之機率並不相等（請參見EPSEM），而是按規模調整其機率比例—由被選為次樣本的單位的數字加以測量。
- 普遍運用於許多的大規模調查方案裡。



6.7 多階段集群抽樣

□ 6.7.4 不等比例抽樣與加權

□ 加權(self-weighting)

- 對不同的個案賦與不同的權重，意即被選為樣本的個案有不同的選取機率。
 - 用最簡單的概念來看，每個個案的權重等於其被選取機率的倒數。當所有個案被選取的機率相同即無需加權。
- 當我們引用不等比例抽樣時，如果沒有伴隨適切的加權措施，將可能發生若干問題。



6.8 機率抽樣的回顧

- 機率抽樣仍然是選取研究元素最有效的方法，因為：
 - (1) 它允許研究者在元素選取過程中避免產生偏誤
 - (2) 它允許誤差的估計



6.9 抽樣的倫理

- 機率抽樣仍存有誤差的風險
 - 研究者必須告知讀者任何可能使結果產生誤導的誤差。
- 當非機率抽樣方法被用來獲得母體中變異的範圍時，研究者必須小心地不要誤導讀者將變異與母體中的典型相互混淆。



The End

