

抽樣與抽樣設計

1

一、抽樣

- 定義：抽樣是指在研究問題確定後，研究者決定要觀察誰、觀察什麼，以及決定由誰回答問題的過程。
- 特質
 - 從母全體中抽取出一部份而已
 - 能代表母全體。
 - 如果抽樣是隨機的且樣本數夠大，大樣本的抽樣可以中和母體中的特殊樣本。

母體

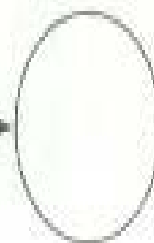
你想要
談的事



抽樣過程

你在資料中真
正觀察到的

樣本



- **母體 (population)**：我們求取資訊的對象全體，可能是人、動物或事物。
- **個體 (unit)**：母體中的一份子。若母體包含的是人，我們則常稱這些人為受試（或受訪）對象 (subject)。
- **樣本 (sample)**：母體的一部份，我們蒐集其資訊以便對整個母體做某些結論。
- **抽樣底冊 (sampling frame)**：個體的清單，我們從抽樣底冊中抽取樣本
- **變數 (variable)**：個體的某種特質，被選入樣本的個體就會被度量這種特質。

- **參數(parameter)**：參數是描述母體的數字。參數是一個固定數字，但我們實際上無法知道參數的值。
- **統計量(statistic)**：統計量是描述樣本的數字。一旦取了樣本，統計量的值就知道了，但是換個不同的樣本，統計量的值就可能改變。我們常用統計量來估計未知的參數。

二、母體

- 抽樣的單位(sampling unit)：母群體內的單一成員。
- 有限與無限母體：取決於抽樣單位是有限或無限的。
 - 有限母體(finite population)
 - 無限母體(infinite population)
- 抽樣的架構(sampling frame)：選擇樣本時，其所涉及實際程序包括了對抽樣單位的完整清單。

三、抽樣設計

隨機抽樣與非隨機抽樣

- 隨機抽樣(probability sample)：母體中每一個抽樣單位，包括樣本內的機率完全相同。
- 非隨機抽樣(nonprobability sampling)：沒有辦法計算包含在樣本中每一個單位的中選機率，而且也不能保證每一個單位都有相同機會被包含在樣本內。

四、隨機抽樣

➤ 選擇機率抽樣的原因與優點

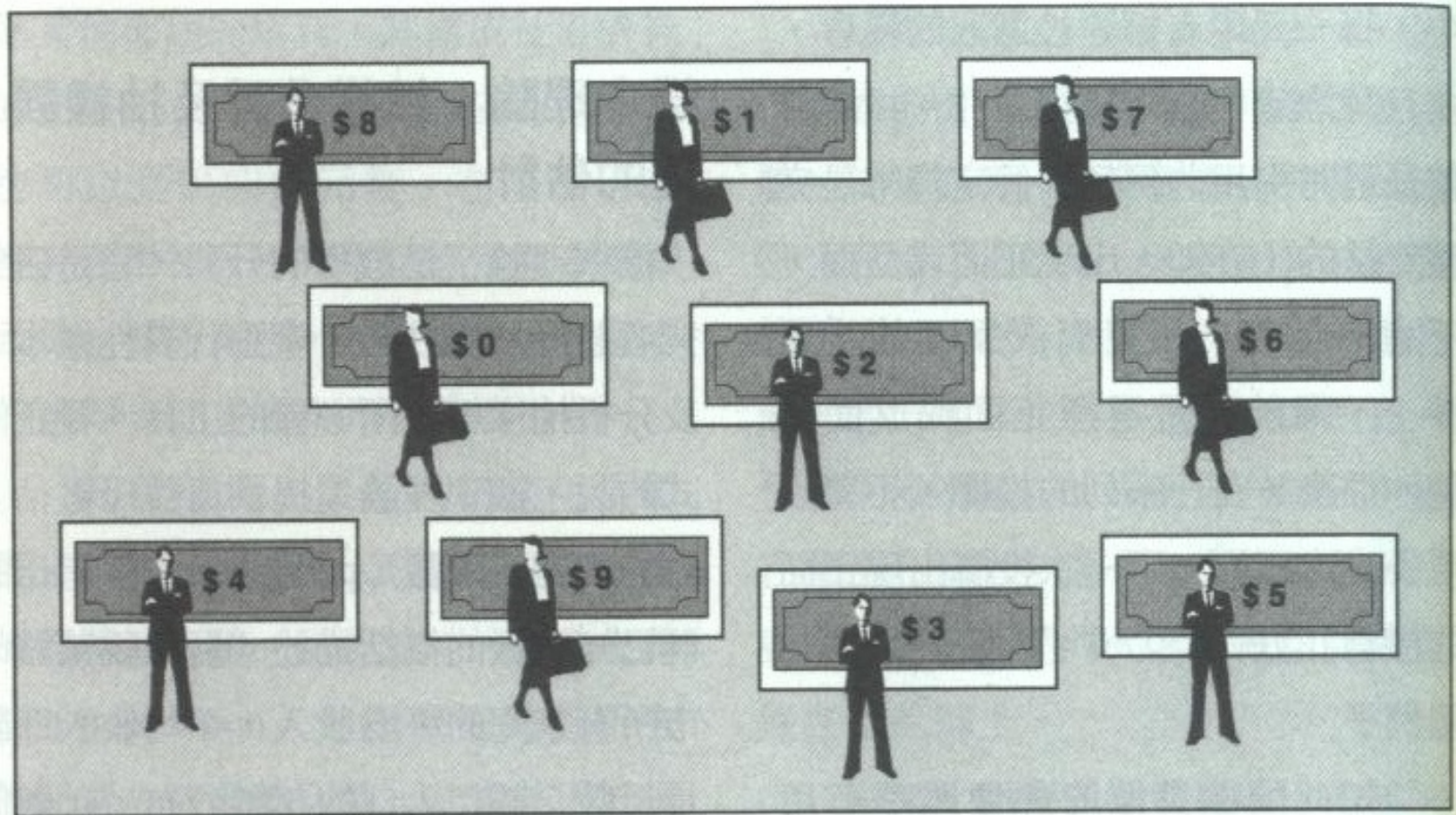
- 1.雖然很少有完美的代表性，機率抽樣所抽取的樣本，一向比其他樣本更具備代表性(representativeness)。代表性因機率樣本而增加，並得以概推以及推論統計之用。
- 2.機率理論得以估計樣本的精確度或代表性。

- 隨機抽樣的類型
 - 簡單隨機樣本
 - 系統隨機樣本(systematic sampling)
 - 分層隨機樣本(stratified sampling)
 - 集群隨機樣本(cluster sampling)
 - 機率隨樣本大小成比例的抽樣(probability proportionate to size, PPS)

簡單隨機樣本(SAMPLE SAMPLING)

- 大小為 n 的簡單隨機樣本是有 n 個個體的樣本，其選取的方法是：使得抽樣底冊任一組 n 個個體，被選中的機率都相同（我們會一直用 n 來代表樣本中的個體個數）。具備這種特質的樣本稱之為相等機率選擇方法（Equal Probability of Selection Method, EPSEM）
- 用兩個步驟選取簡單隨機樣本
 - 步驟1：編代碼。對抽樣底冊中每個個體指定一個數字代碼。
 - 步驟2：用亂數表。利用隨機數字來隨機選取代碼。

擁有\$0-\$9 的 10 人母群

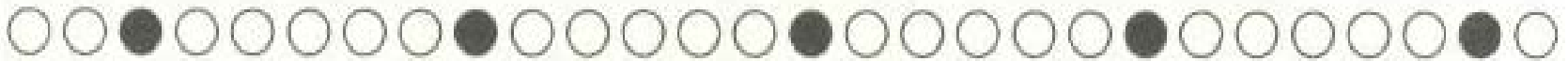


例：從850名學生中抽出10人接受測驗，如何從亂數表中選取？

10819	85717	64540	95692	44985	88504	50298	20830	67124	20557
28459	13687	50699	62110	49307	84465	66518	08290	96957	45050
19105	52686	51336	53101	81842	20323	71091	78598	60969	74898
35376	72734	13951	27528	36140	42195	25942	70835	45825	49277
93818	84972	66048	83361	56465	65449	87748	95405	98712	97183
35859	82675	87301	71211	78007	99316	25591	63995	40577	78894
66241	89679	04843	96407	01970	06913	19259	72929	82868	50457
44222	37633	85262	65308	03252	36770	51640	18333	33971	49352
54966	75662	80544	48943	87983	62759	55698	41068	35558	60870
43351	15285	38157	45261	50114	35934	05950	11735	51769	07389
11208	80818	78325	14807	19325	41500	01263	09211	56005	44250
71379	53517	15553	04774	63452	50294	06332	69926	20592	06305
63162	41154	78345	23645	74235	72054	84152	27889	76881	58652
17457	68490	19878	04981	83667	00053	12003	84614	14842	29462

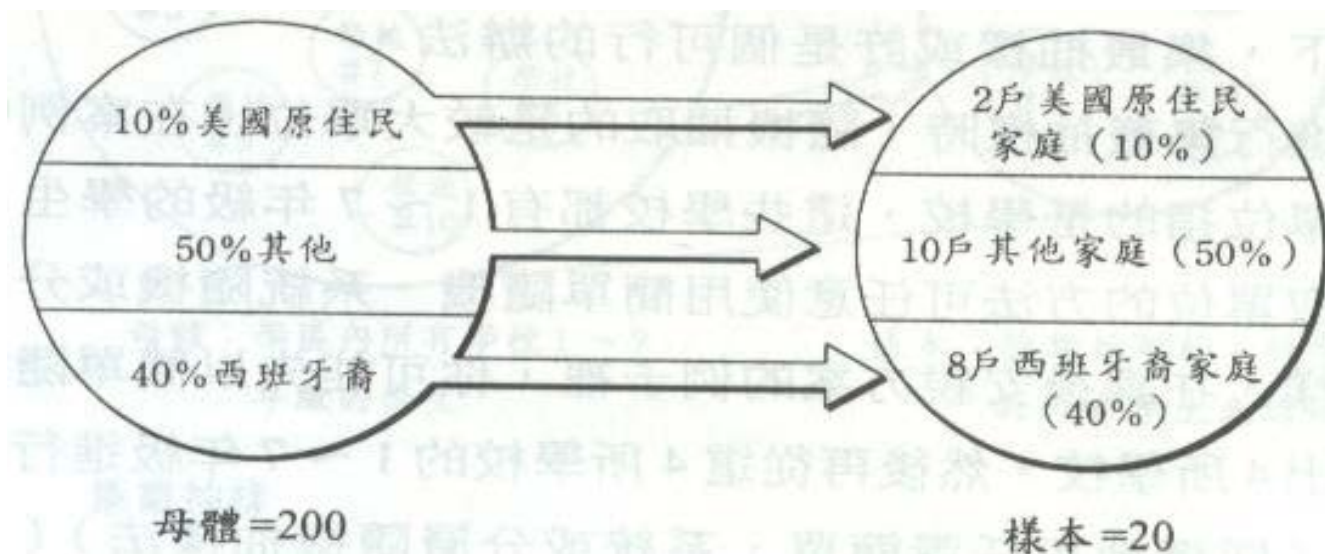
系統隨機樣本(SYSTEMATIC SAMPLING)

- 計算母體總數目，根據所需的樣本數，在母體中每隔一定的距離（抽樣間距=母群大小/樣本數）抽取出樣本，依隨機原則開始選取。
- 如從900個學生中抽出150個，開始先隨機挑出第一個，然後每隔6個就抽一個學生，抽樣間距是6（如下圖）



分層隨機樣本(stratified sampling)

- 先將母體依明確特徵分成若干層，每一層內之個體特質應盡量相同，而與其他層特質應盡量相異。
- 計算母體中每一層個體所佔比例，然後依此比例分配樣本數，使各層樣本所佔比例與母體比例完全相同。
- 依簡單隨機抽樣法、亂數表或系統抽樣法，抽出需要的樣本數。



系統分層抽樣舉例

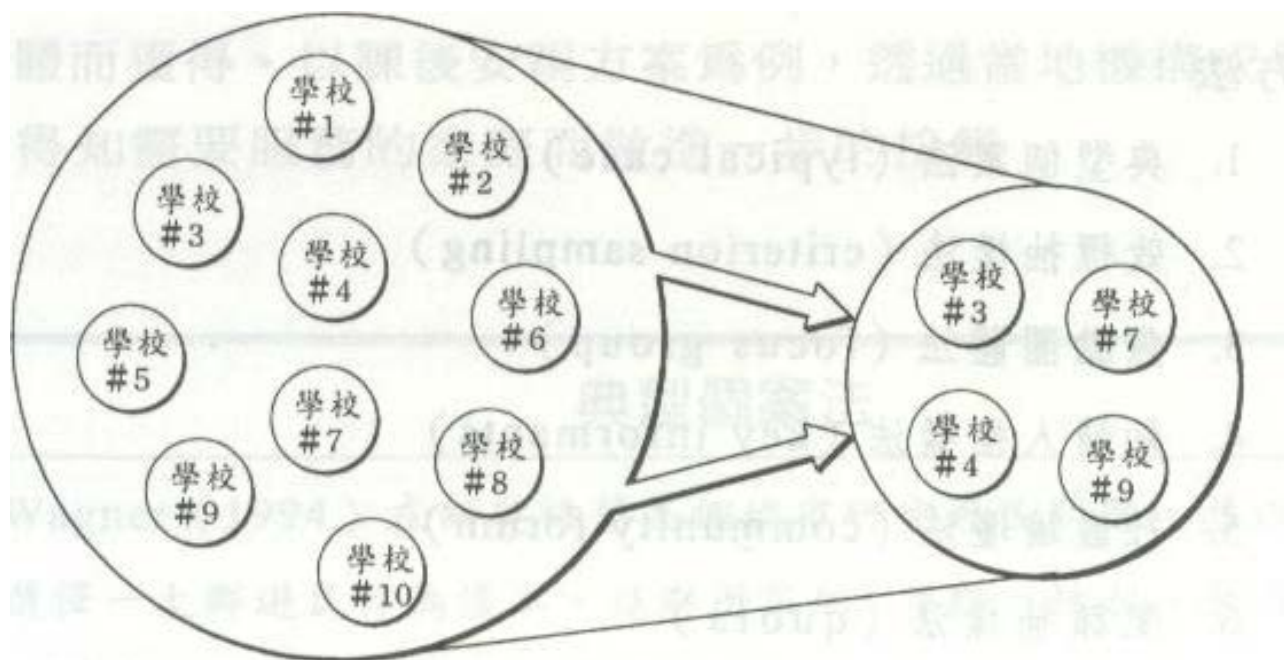
訪問某校765位學生。先抽出5個學院，再由每個抽中的學院中訪問153人（ $5 \times 153 = 765$ ），抽樣間距1880（ $9404/5$ ），亂數起始1203

學院	學生數	累積人數	學生編號	抽出學院	學院學生樣本數
教育	233	233	1-233		
國際事務	274	507	234-507		
理	616	1123	508-1123		
法	642	1765	1124-1765	1203	153
傳播	791	2556	1766-2556		
文	835	3391	2557-3391	3083	153
外語	1118	4509	3392-4509		
社科	2281	6790	4510-6790	4963	153
商	2614	9404	6791-9404	4843,8723	306
總數	9404				765

集群隨機樣本(cluster sampling)

- 以群體而非個人為抽樣單位，進行抽樣時，要先明確劃分出數各群體，各群體間應是同質，群體內成員則為高異質性，由於群體間差異通常頗大，故群體數目不能太少，以減少抽樣誤差，而群體內成員數目不宜過多，以利全體調查的可能進行，劃分出數各群體後，再用簡單隨機方式，抽出其中的幾個群體，然後訪問群體內的所有個體

例：研究國小課後安親方案之需求



母體：學區內所有學校1～7
年級的學生

樣本：隨機抽取的4所學校
的所有學生

機率隨樣本大小成比例的抽樣(probability proportionate to size, PPS)

- 屬於多段集群隨機抽樣的一種。每一集群樣本的選取不是相等的機率，而是隨著樣本大小成比例的抽樣，依照次樣本數計算。

機率抽樣

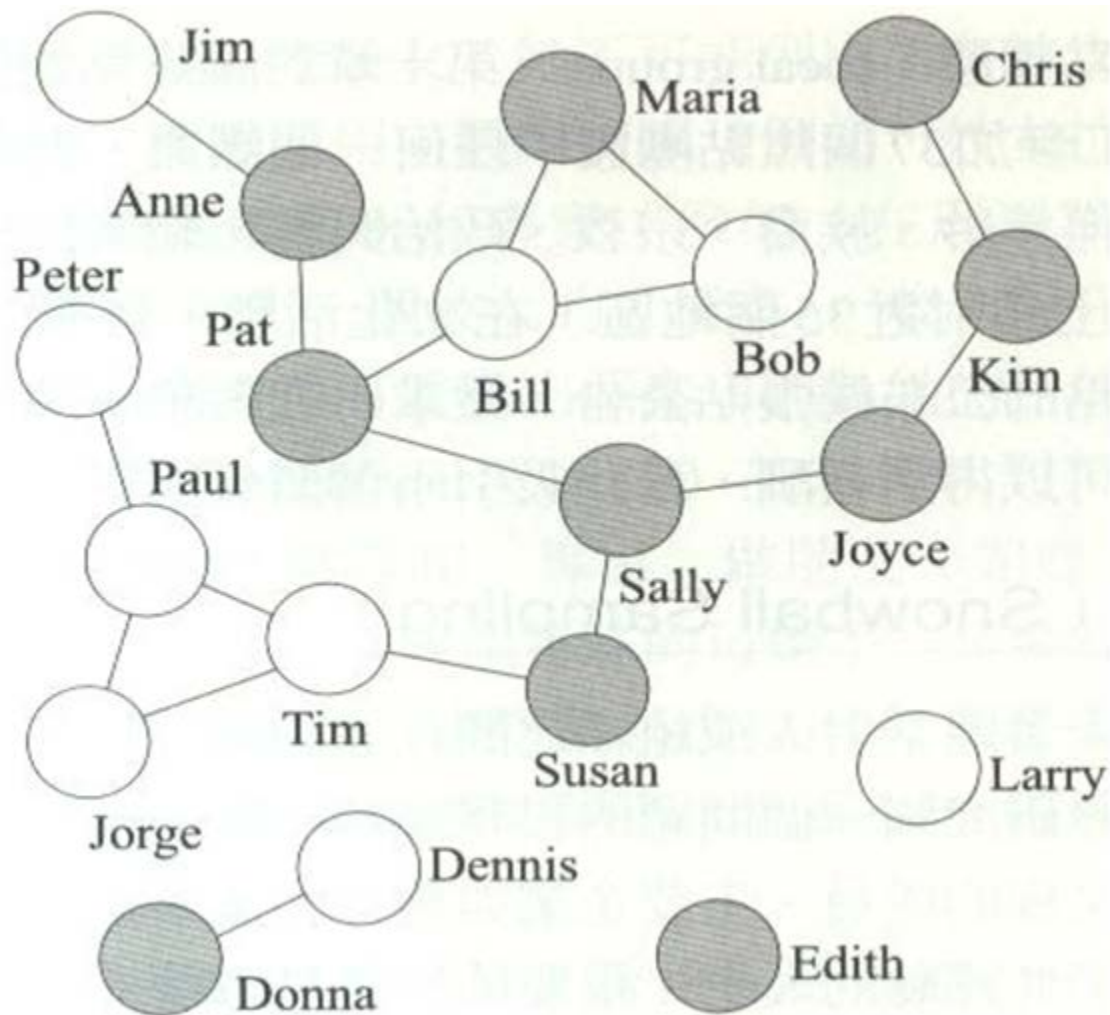
類型	敘述	優點	缺點
簡單隨機抽樣	母體中每個元素被抽出的機率相同。實施的方法有號碼球取樣法及亂數表法。	1.於學理上最精確 2.僅受隨機誤差影響。	1.母體若很大時不易實施。 2.需要母體全部的名冊方能實行。
系統抽樣	將抽樣架構中各元素依次編號分組，選取架構中第 k 個元素組成的樣本。	1.較隨機抽樣容易實施、節省成本。	1.若抽樣架構具有週期性則樣本可能不具代表性。
分層抽樣	為前兩種抽樣法的混合體。先將母體分成幾個同質的次集合，再從其中抽出樣本。	1.比前二者更具抽樣效率，研究者若想研究次母體特質時，是一個不錯的方法。	1.正確的分層不是件很容易的事，可能要花費許多成本。
群集抽樣	先將母體分成若干群，再以隨機的方式抽出若干群之中的數群，對群內的樣本全數訪問。	1.節省研究時間與財力。 2.抽樣架構中每個元素資料不易得到或不完整時，使用此法會較方便。	1.群集的大小差異會影響抽樣正確性。 2.依目的可能要抽樣二次以上，例如先抽村里，再抽戶。發生抽樣錯誤的機率較高。



五、非隨機樣本的設計

- 權宜(便利)樣本(convenience sample)：研究者選擇非常容易得到的抽樣單位。
- 目的(立意)樣本(purposive sample)：研究者主觀選擇一個抽樣單位，來獲得一個母體的代表性樣本。
- 配額樣本(quota sample)：以人為方式儘可能選定一個與抽樣母體相似的樣本。
- 滾雪球抽樣

滾雪球抽樣



非機率抽樣

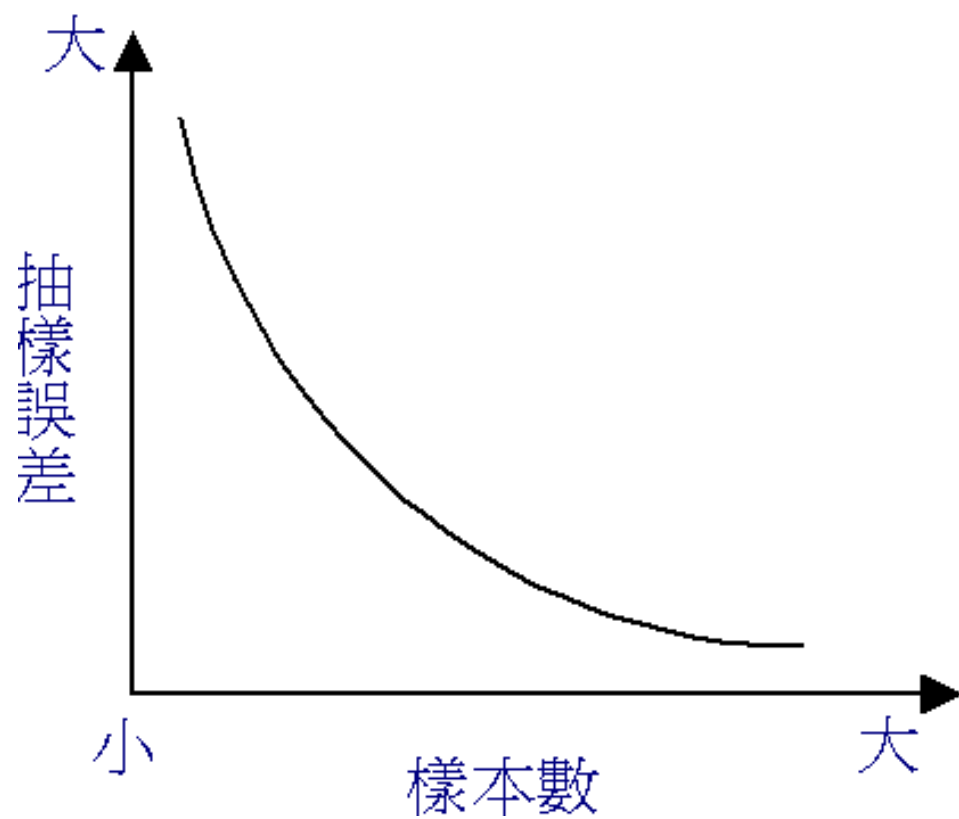
類型	敘述	優點	缺點
便利抽樣	研究者使用最方便或是最經濟的方法來進行抽樣。	1. 不需要母體的名冊。 2. 快速、便利。	1. 正確性和估計偏差不能衡量或控制。 2. 研究者的主觀意識可能影響抽樣，選出的樣本可能不是很適合代表母體。
配額抽樣	研究者將母體依特質區分為數類，而抽樣時按比例從各類中抽出；其樣本元素具有某種特質的比率和母體元素具有某種特質的比率大約是一致的。	1. 較機率抽樣中的分層抽樣成本低。 2. 具有分層抽樣的效果。	1. 雖採用配額的方式抽樣，但在抽樣時若不是隨機選取，選出的樣本也會有誤差，而不能代表整個母體。 2. 在研究者將母體分類時可能會產生偏誤。
判斷抽樣	依研究者的判斷進行抽樣，研究者對母體必須有深入的了解。	1. 在某種類型如選舉預測上是很有用的。 2. 在蒐集樣本時，較節省成本及時間。	1. 研究者在抽樣時可能會因主觀因素而影響了抽樣，造成偏差。 2. 由抽樣資料來推估母體時較不適合。
雪球抽樣	先蒐集目標母體的少數成員，再由這些成員引出其他的母體成員。	1. 在尋找少數難以尋找的母體時，此法是一個很好的方法。	1. 因為抽樣單位不獨立，會產生較高的偏差。 2. 由抽樣資料來推估母體時較不適合。

抽樣誤差與非抽樣誤差

- 抽樣誤差：抽樣過程中所產生的誤差，原因可能是抽樣方法不對或設計不當
- 非抽樣誤差：蒐集或處理資料過程中產生的誤差，如反應錯誤、偏見、拒訪、抄寫錯誤、對問題不清楚等



抽樣誤差與樣本數

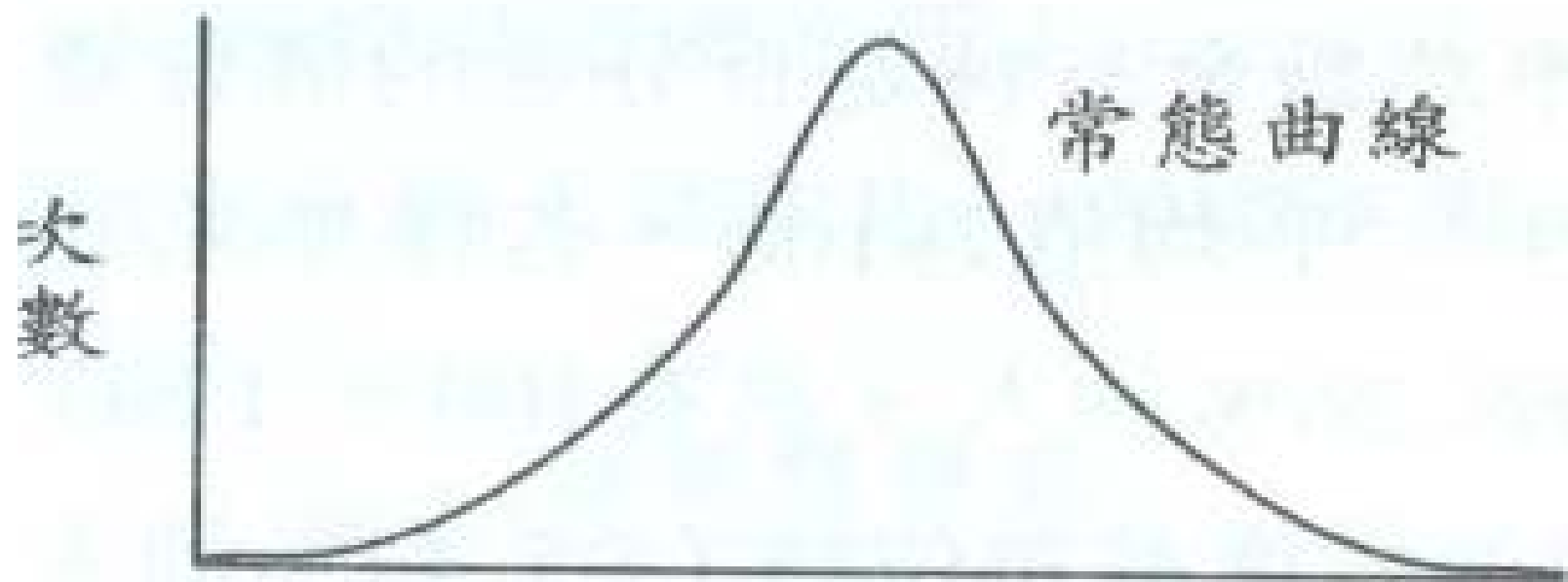


六、樣本規模

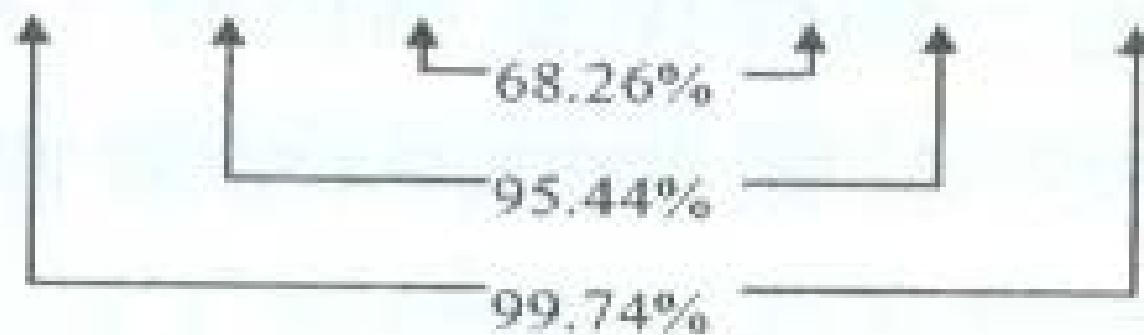
- 標準誤(standard error)：用來顯現樣本結果反映真正參數值接近程度的統計測量。
- 信賴區間(confidence interval)
- 樣本規模的決定

信賴敘述

- 信賴敘述包含兩個部分：信賴區間(confidence interval)及信賴水準(confidence level)。
- 信賴區間告訴我們樣本統計量離母體參數多遠，也就是樣本統計量的精確程度。
- 信賴水準告訴我們所有可能樣本中有多少百分比滿足這樣的信賴區間，也就是指統計量會落在相距母數的某一特定區間內的信心程度。



-3SD -2SD -1SD 0 +1SD +2SD +3SD



信賴區間

○ 【99年社會工作師專業職業及技術人員考試】

社會工作領域運用質性研究法時，如何決定抽樣策略及樣本的大小規模？試舉例說明之。

想一想～～

- 1.在行為科學中我們通常把焦慮或是沮喪的量表視為何種尺度？(A)名目尺度；(B)次序尺度；(C)等距尺度；(D)等比尺度。
- 2.家中的經濟狀況會透過個人的教育水準影響個人未來的生涯發展，請問個人的教育水準是什麼變項？(A)依變項；(B)自變項；(C)抑制變項；(D)中介變項。
- 3.電視節目中，民眾的叩應節目是何種抽樣？(A)隨機抽樣；(B)非隨機抽樣；(C)系統抽樣；(D)立意抽樣。
- 4.當樣本蒐集不易，所以請既有的樣本介紹受訪者，再藉此擴散出去，此種抽樣方法為：(A)隨機抽樣；(B)滾雪球抽樣；(C)立意抽樣；(D)系統抽樣。



- 5.抽樣的主要原則為何？(A)最小成本原則；(B)最具代表性原則；(C)抽取最多數樣本原則；(D)最具特殊性原則。
- 6.請問以下何者不屬於一客觀公正的民意調查應符合的原則？(A)隨機抽樣；(B)樣本數要夠大；(C)將民意調查視為影響選民的工具；(D)問題的設計不具引導性。
- 7.以下何者不是文獻回顧的目的？(A)了解到目前為止別人的研究發現；(B)找到本研究的理論依據；(C)找到可能支持研究經費的機構；(D)蒐集本研究可以使用的研究方法。
- 8.以下何者為好的文獻整理？(A)只蒐集和結果一致的文獻；(B)將別人的文獻資料整理的愈完整愈好；(C)將蒐集的資料和研究結果融會貫通，融入自己的研究之中；(D)以上皆非。



- 1.B
- 2.D
- 3.B
- 4.B
- 5.B
- 6.C
- 7.C
- 8.C

