

抽樣和隨機抽樣

抽樣(sampling)是指從母群體中抽取一部分個體作為樣本，通過觀察樣本的某些屬性達到對母群體的認識。

如果母群體中的每個個體都有一個已知的、非零的機率被選為樣本，雖然被選為樣本的機率不一定要相同，但必需知道其被選的機率，就可以稱為隨機抽樣(Random Sampling)，在大多數狀況之下，母群體中的每個個體被抽中的機率是相同的，所以隨機抽樣也就是俗話說的「公平」！由於隨機抽樣並沒有特別要去抽取任何個體的傾向，因此可以避免抽樣上的偏差，它有健全的統計理論基礎，可以估計樣本統計量的誤差或可靠性。

隨機抽樣看似簡單易懂，但是很多同學在學習時並沒有真正理解。隨機抽樣也有簡單隨機抽樣法、分層隨機抽樣法、集群隨機抽樣法、系統抽樣法等多種方法，我們會在另文中詳細說明。但不管用哪一種方法，都是有一定步驟的，並不是我們以為的就是「任意」去抽，也就是說，所謂的「隨機」並不是「隨意」。

例如老師上課時要隨機的抽二位同學來回答問題。他說，因為今天是 30 號星期 4，那就 30 號和 4 號吧！請問，這是隨機抽樣嗎？不是！這只是隨意的抽樣，統計術語稱為「非隨機抽樣的便利抽樣法」。為什麼？因為這是以班上所有同學為母群體，如果用這種方法來抽的話，每一個個體被抽中的機率是不相同的，用俗話講就是並不公平。為什麼不公平？因為一個月最多 31 天，那 32 號以後的同學鐵定不會被叫到，又一個星期只有 7 天，只有座號個位數是 1 至 7 的才會被叫到，這是任意指定，當然不是隨機抽樣！所以以後我們只要觀察看看，是不是母群體中的每個個體都有相同的機率被抽到，只要能舉出一個不公平的例子，這個方法就不符隨機的要求。同學想想看，如果是隨便翻教科書的頁碼，可以算是隨機嗎？

要真正做到「隨機」，其實並不容易，有些情況，表面看起來好像很公平，仔細分析起來，也是有問題有漏洞的。美國在 1940 年的徵兵抽籤為了確保公平、公開且隨機的從一個大箱子裡選出內含數字的膠囊，膠囊中的數字則對應至役男的編號，結果卻發現嚴重的偏離真正的隨機設計。原來是這些膠囊在抽選前並沒有好好的混合均勻，因為要混合均勻這麼多的膠囊事實上是「非常非常困難」的！

1970 年美國政府抽籤徵兵時為了避免 1940 年的瑕疵，就做了慎重的改

進。這次抽籤是根據生日，一年中的每一天對應至 1 到 366(包括潤年的 2/29)。他們的做法如下：首先將 1 月的 31 個日期寫在紙條上，然後把紙條分別塞進 31 個膠囊內再放到一個眼大的矩形箱子內。然後再把 2 月的 29 個日期寫在紙條上、塞進 29 個膠囊內後又置入箱中與一月的 31 個膠囊混合均勻，如此重覆到 12 月的 31 個投入後混合均勻。在抽籤時，第一個抽出的膠囊編上 1 號，第二個編 2 號，以此類推，直到所有的 366 個膠囊都抽出為止。然後按徵兵需求名額依抽籤的序號開徵。請同學們仔細想想，這樣是否隨機？很不幸的，1970 年的抽籤同樣並未通過隨機的統計檢定，因為如果將這 366 個數字分成三組：1~122、123~244、245~366，結果根據實際抽出的結果，發現在這三組內，各月份的分布情況並不相同，例如在第一組中，一到四月份出現的次數明顯少於其他八個月份，十二月份的 31 個竟有 17 個是在 1~122 這一組，但一月份的只有 9 個。好了，同學們請仔細想想，你能夠指出問題出在哪裡嗎？

(解答)

- 膠囊混合得並不均勻

看起來膠囊好像都混了很多次應該會均勻，實際上是一月份和其他月份的膠囊總共混合了 11 次，二月分的混了 10 次……，十二月份的只混合了一次。所以一月份的分佈會比較平均，十二月份的偏在上層附近位置的最多，十一月的次之，依此類推。而且箱中的膠囊愈來愈多，有好幾十幾百個膠囊時實務上更難充分混合均勻。

- 抽取時取樣點變化不大

在抽順序時，負責抽籤的人大部份的時候似乎都只是從箱子的上層中抽出來，取樣點的位置變化不大，幾乎沒有人會刻意的把手伸到箱子的非常下層或旁邊來抽。

思考練習：

- 我們在學生餐廳吃飯時間，以桌號為準隨機抽取樣本發放問卷，請問是否隨機抽樣？為什麼？
- 我們找一個人多的路口發放問卷，算不算隨機抽樣？為什麼？

- 在學校公告欄公開徵求預定數目的自願者，參與本研究的實驗(當然需付給酬勞)，請問這組樣本是否來自隨機抽樣？為什麼？
- 設計好問卷後，以本專題組成員為種子，請個自利用電腦轉發自己的朋友圈(儘量避免重複)，再請朋友幫忙轉發他們的朋友圈填卷，以達到預定的回收份數為目標。請問這是不是隨機抽樣？為什麼？