

單元 12: 隨機抽樣

(課本 §2.12)

抽樣的方法會影響某一特別樣本結果的機率，如由某一含有 $N = 5$ 個元素的母體中，選出大小為 $n = 2$ 的一樣本，兩種可能的方法為

- (1) 不放回取樣 (sampling without replacement): 徹底混合後，一次選出 2 個使得每一對元素被選中的機率都一樣，則

$$P(\text{選出某一特定對元素}) = \frac{1}{\binom{5}{2}} = \frac{1}{10}$$

或選出一個後，不放回，再選另一個，得

$$P(\text{選出某一特定對元素}) = \frac{1}{5} \cdot \frac{1}{4} \cdot 2 = \frac{1}{10}$$

- (2) 放回取樣 (sampling with replacement): 選出一元素後，放回，再選出一元素，則

$$\begin{aligned} P(\text{選出與 (1) 中相同的特定對元素}) \\ = 2 \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{2}{25} \end{aligned}$$

因此，需要清楚地敘述抽樣過程，以便能有效地由樣本對母體做推論。

爲實用計，僅介紹如下定義的簡單隨機抽樣 (simple random sampling)，簡稱隨機抽樣 (random sampling):

定義 1. 令 N 爲母體的大小，且 n 爲樣本的大小。若抽樣的執行程序會使得所有 $\binom{N}{n}$ 個樣本中，每一樣本被選出的機會均等的話，則稱此種抽樣爲隨機抽樣 (random sampling)，得到的結果爲一隨機樣本 (random sample)。

註 1. 執行隨機抽樣的方法：

- (1) 母體不大時：標號在 N 個相同的球（或牌）上，均勻混合後，選出 n 個球（或牌），再找出對應的度量值。
- (2) 亂數表 (Table of Random Numbers): 乃由一組數字 $(0, 1, \dots, 9)$ 所生成，以致於在長期生成下，所有 10 個數字在表中出現的比率近似相等，

且沒有任何趨勢或樣式的顯現，使用方法說明如下：
設由 $\{1, 2, \dots, 7\}$ 中任選三數，過程為

- (i) 確定起點：任選一處，如，第 9 行，15 列，最右邊之數：5 (第 1 個)
- (ii) 確定方向：任選一個，如，下：2 (第 2 個)
- (iii) 持續 (ii) 的方向，選出適當的數，略過不適當的數，直到選出所要的大小，如，8 (不合，超過 7)，5 (不合，重複)，5 (不合，重複)，1 (第 3 個)。故得樣本 $= \{5, 2, 1\}$ 。

註 2. 當母體是觀念上的，如實驗室的實驗結果，則以重複實驗的方式，獲取所需大小的樣本，以作為一隨機樣本的合理程度的近似。

註 3. 某些抽樣技術是刻意地設計成部分地隨機，以便能獲得具代表性的樣本，如一樣本中的選民，若由隨機抽樣方式產生，則可能由於純機會的緣故，會全是加州的選民，而喪失代表性，故先分派每一州特定數量的樣本大小，而選出每一州的樣本（稱作子樣本），再合併成全國性的整體樣本，則較具有代表性。

註 4. 一般而言, 期望以最小的成本選出含有特定數量訊息的樣本.