分配基本觀念

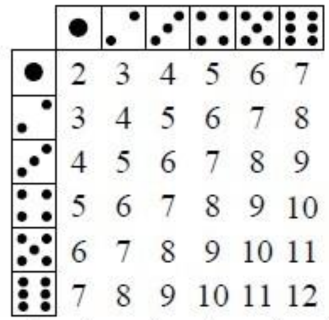
「分配」是何物？

機率、機率模型、機率分配

事件：投擲三顆骰子，總數出現18的機率？

算出擲出 18 為止的次數
「6」、「6」、「6」

算出機率為 1/216

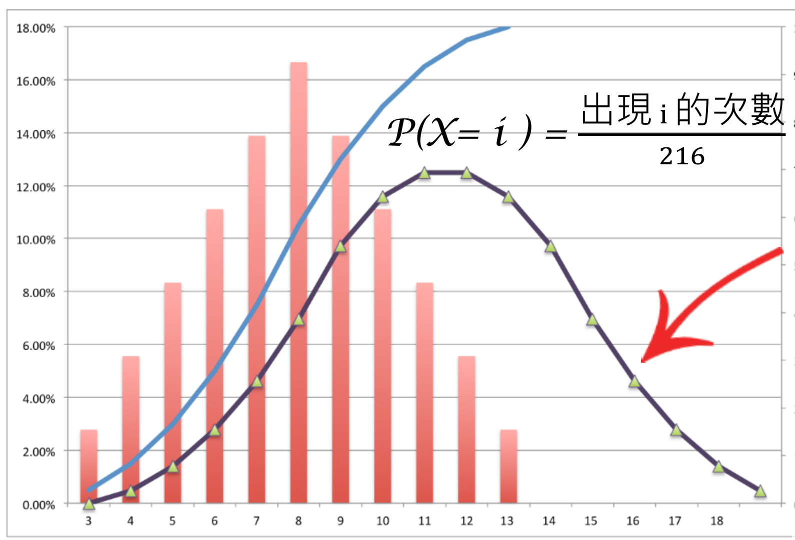


總和	2	3	4	5	6	7	8	9	10	11	12
次數	1	2	3	4	5	6	5	4	3	2	1

另 X 表示隨機變數

$\mathcal{P}(X=18) = 1/216$

$$\mathcal{P}(X=i) = \frac{\text{出現 i 的次數}}{216}$$

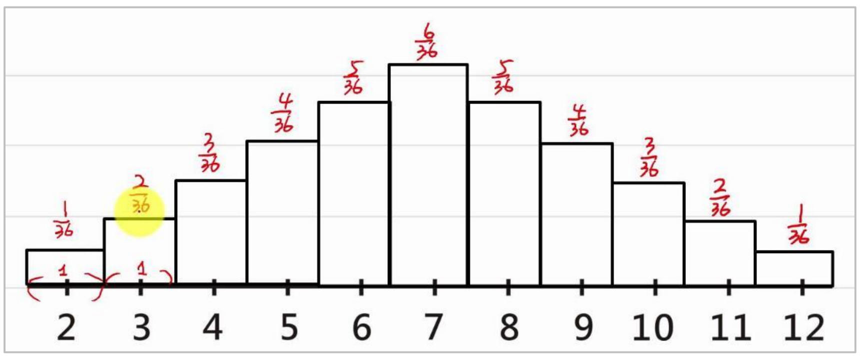


如果某事件可以把所有發生的機率呈現出來（例如圖形），則所呈現的型態、特質就是所謂該事件的「分配」

機率分配

機率分配的表示方式一般有<u>機率分配圖</u>、<u>機率分配表</u>與<u>機率函數</u>三種方式

機率分配圖



機率分配表

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

總和	2	3	4	5	6	7	8	9	10	11	12
次數	1	2	3	4	5	6	5	4	3	2	1

機率函數

$$f(x)=\begin{cases}\dfrac{1}{36} & ,x=2,12\\ \dfrac{2}{36} & ,x=3,11\\ \dfrac{3}{36} & ,x=4,10\\ \dfrac{4}{36} & ,x=5,9\\ \dfrac{5}{36} & ,x=6.\ 8\\ \dfrac{6}{36} & ,x=7\end{cases}$$

常見離散型機率分配

分配類型	適用情境說明	分配函數
均勻分配 discrete uniform distribution	無論隨機變數如何變化，其機率值為固定常數，例如投擲一次骰子，機率都是1/6	$f(x) = \begin{cases} \frac{1}{n}, x = 1,2,..,n \\ 0, \text{other} \end{cases}$
二項分配 binomial distribution	在成功與失敗的機率不變情況下，不斷重複試驗所得的機率分配函數	$f(x) = C_x^n\, p^x\, q^{n-x}$, x=0,1,..,n 其中p為成功機率；q為失敗機率
伯努利分配 bernoulli distribution	在二項分配下，不論成功或失敗，只做一次（n=1），算是二項分配的特例	$f(x) = p^x\, q^{1-x}$, x=0,1 其中p為成功機率；q為失敗機率
超幾何分配 hyper geometric distribution	從N物中採抽出不放回抽樣，抽出n個樣本，其中N物中，有S個相同，N-S個不同。所以每次抽樣都會受前次抽樣影響	$f(x) = \frac{C_x^S\, C_{n-x}^{N-S}}{C_n^N}$, $0 \le x \le \min(n,S)$
波松分配 Poisson distribution	在一個單位時段或區段內，某事件發生次數（X）的問題。	$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0,1,2 \ldots$ λ為平均次數

其他離散型機率分配

分配類型	適用情境說明	分配函數
多項分配 multinomial distribution	二項分配的延伸，二項分配是把母體分成二種類別，而多項分配則將母體分類超過兩個以上	$f(x_1, x_2, \ldots, x_k)$ $= \dfrac{n!}{x_1!\ x_2!\ \cdots\ x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$ $= \mathrm{C}^{\mathrm{n}}_{x_1,x_2,\ldots,x_k}\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$
負二項分配 negative binomial distribution	第 r 次成功所需試驗的次數 （每次都是伯努利試驗） （X表示第 r 次事件發生時所需試驗次數）	$f(x) = C^{\mathrm{x-1}}_{\mathrm{r-1}}\ \mathrm{p^r}\ \mathrm{q^{x-r}}$, $x = r, r+1, r+2, \ldots..$ 其中p為成功機率；q為失敗機率
幾何分配 geometric distribution	不斷進行伯努利分配，直到第1次成功為止所需試驗次數之機率函數	$f(x) = \mathrm{q^{x-1}p}$, x=0,1,2,.... 其中p為成功機率；q為失敗機率

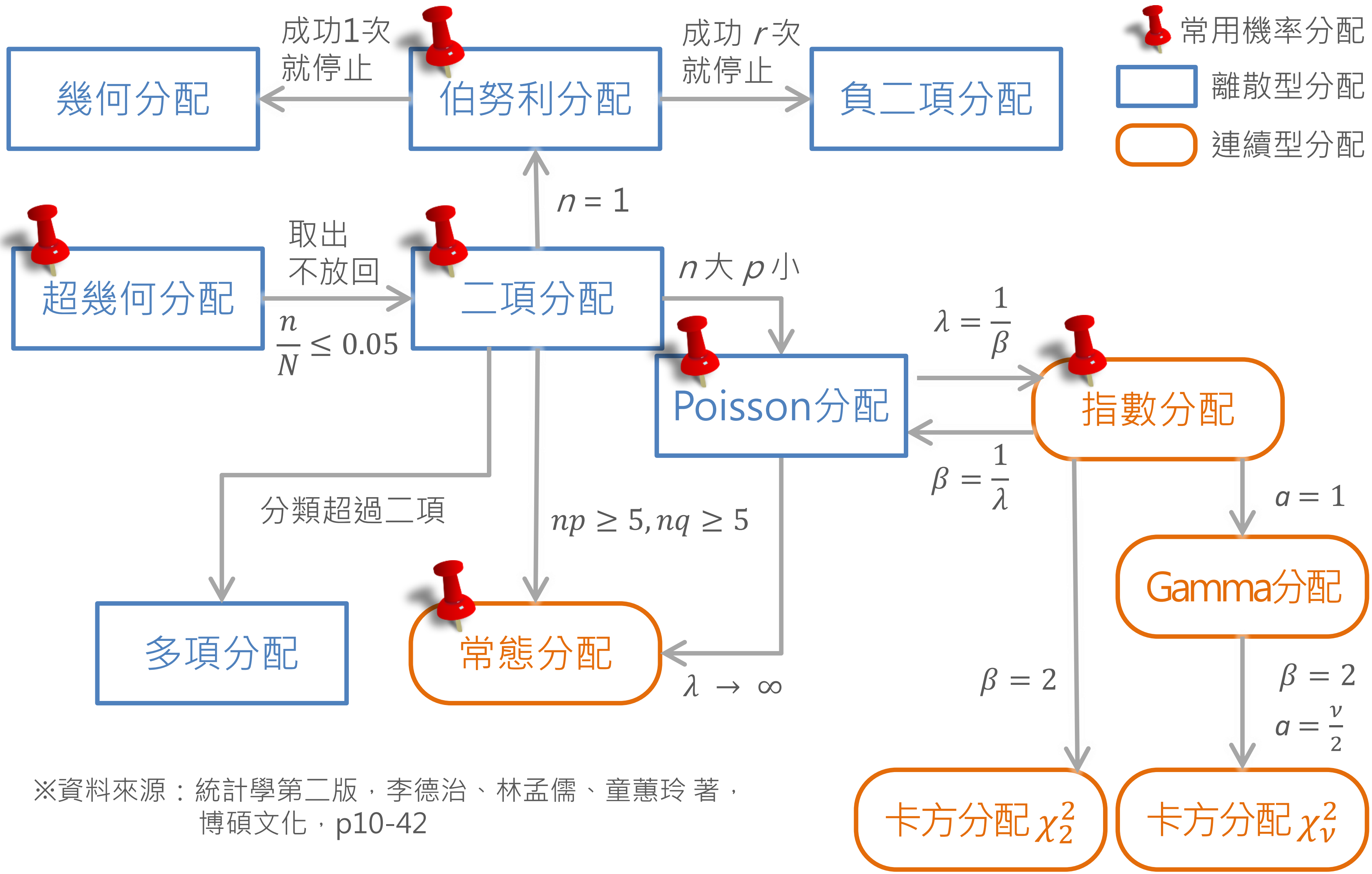
常見連續型機率分配

分配類型	適用情境說明	分配函數
連續均勻分配 uniform distribution	隨機變數(X)在某連續區間內所發生的機率都相同	$f(x) = \frac{1}{b-a}$, $a \leq x \leq b$
常態分配 normal distribution	存在於大自然間的各種現象或狀態，都可以將母體視為常態分配	$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$ $-\infty < x < \infty$, μ為平均數, σ為標準差
標準常態分配 standard normal distribution	透過 $z = \frac{x-\mu}{\sigma}$ 的「標準化」，將原本要利用微積分計算求值，轉換成可以利用查表得到結果	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $-\infty < z < \infty$
指數分配 exponential distribution	與Poisson隨機變數相反，指數分配的隨機變數(X)是描述連續兩事件發生的間隔時間	$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$, $x \geq 0$ x 為第一次發生事件所需時間； β 為事件發生的平均時間；

其他連續型機率分配

分配類型	適用情境說明	分配函數
Gamma 分配 Gamma distribution	隨機變數（*X*）表示事件第 *a* 次發生所需的時間。因此若「*X*>*t*」表示事件第 *a* 次發生至少需要 *t* 個時間單位；另個說法是，在 *t* 時間內事件至少發生 *(a-1)* 次的機率	$f(x) = \frac{1}{\beta\Gamma(a)}(\frac{x}{\beta})^{a-1}\, e^{-\frac{x}{\beta}}$, $0 < x < \infty , a > 0, \beta > 0$ x 為第 *a* 次發生所需時間； β 表發生一次所需的時間；
卡方分配 chi-square distribution	卡方分配可以算是Gamma分配的特例（$a = \frac{\nu}{2}$, $\beta = 2$），在統計應用上，可進行單一母體變異數 σ^2 的統計推論；可用來做適合度檢定（goodness-of-fit test）、獨立性檢定（test of independence）與變異數齊一性檢定（test of homogeneity）	$f(x) = \frac{1}{2\Gamma\left(\frac{\nu}{2}\right)}(\frac{x}{2})^{\frac{\nu}{2}-1}\, e^{-\frac{x}{2}}$, $0 < x$, ν 為卡方分配的自由度

各種機率分配的關係



※資料來源：統計學第二版，李德治、林孟儒、童惠玲 著，博碩文化，p10-42

案例說明

~利用案例來解說觀念

政府調查受僱者薪資

案例介紹

【中國時報 2016/11/8】

行政院主計總處昨日首度發布「薪資中位數報告」，指出去年我國每月薪資（含經常性及非經常性）中位數40853元，占薪資平均數48490元比率為84.25%，創近7年新低，顯示我國受薪階級向低薪集中的情形，有逐年擴大趨勢，值得密切注意。

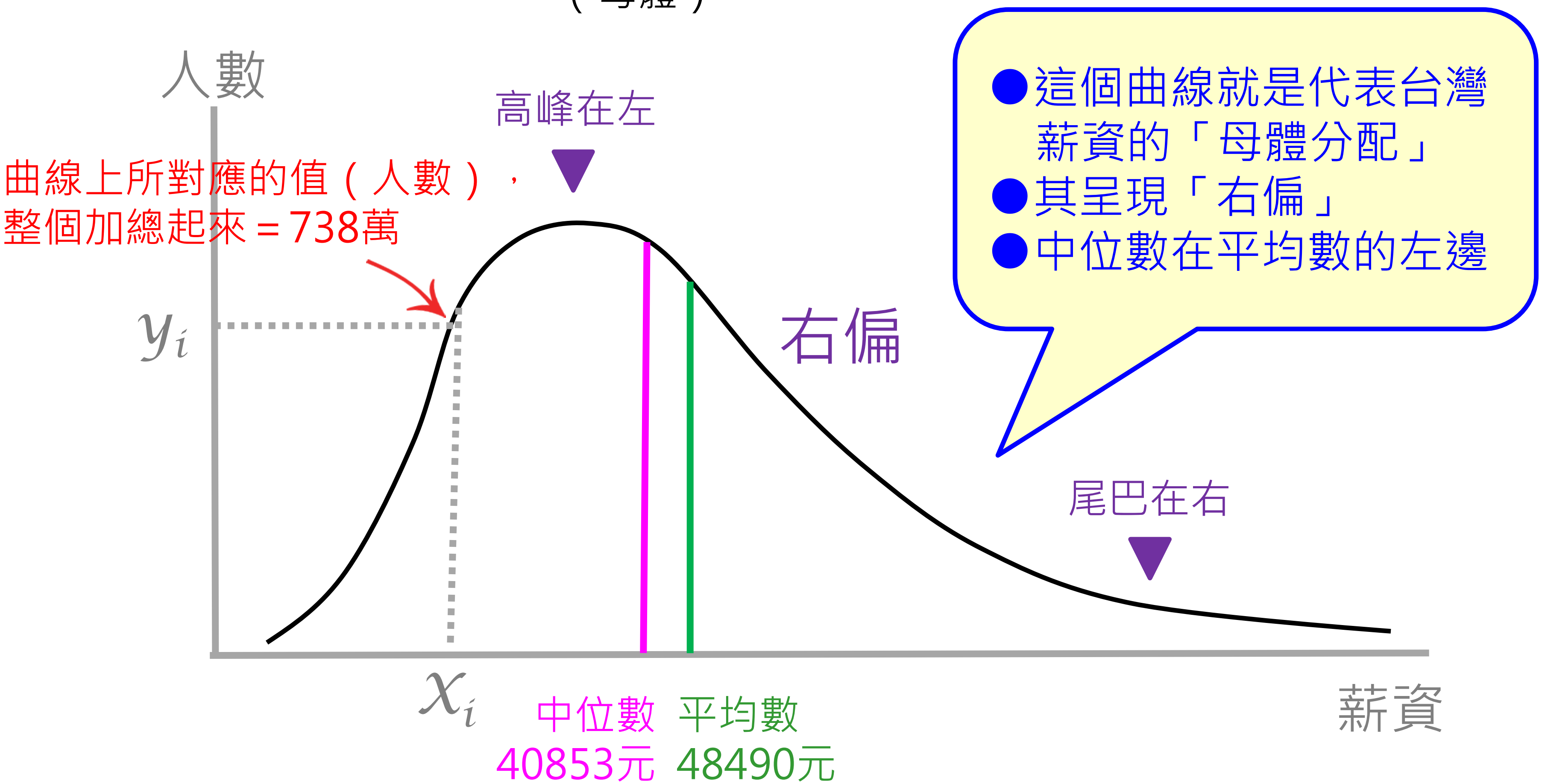
主計總處指出，以薪資為橫座標，以人數為縱座標所繪得的分布圖，近年已呈現高峰在左、尾巴在右的「右偏」型態，反映低薪人數愈來愈多，國人有向低薪集中趨勢，以致薪資中位數明顯低於平均數......

主計總處官員表示，美國薪資中位數占平均數比率是74.90%、日本86.66%、澳洲84.10%，各國薪資分布普遍都有「右偏」現象，尤其美國右偏幅度更大，顯示其高、低薪差距更為嚴重。 主計總處表示，綜合各國統計可以發現，除了薪資，房價、財富的分布也都有「右偏」的情況....

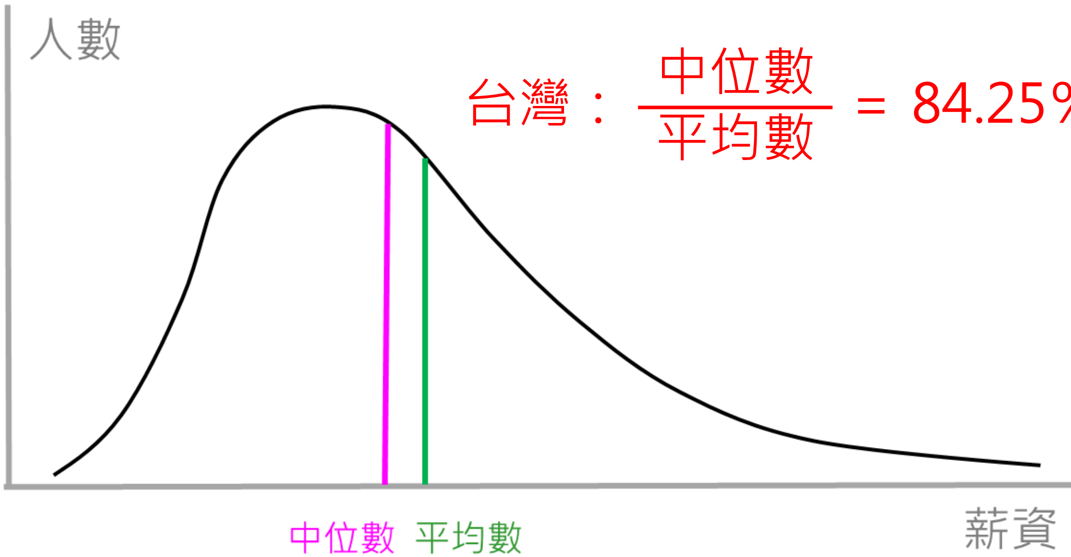
母體分配的案例

假設政府把所調查的全部738萬名受僱者薪資數據整理成下圖

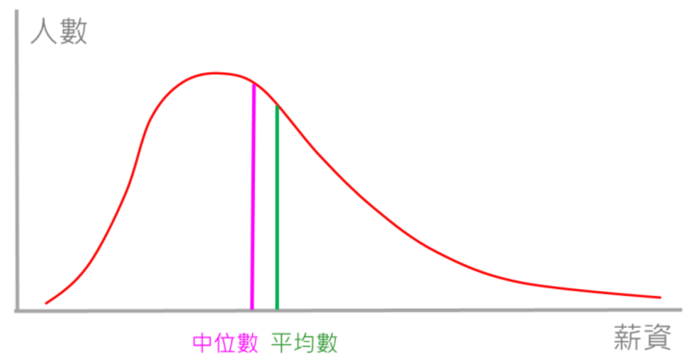
（母體）



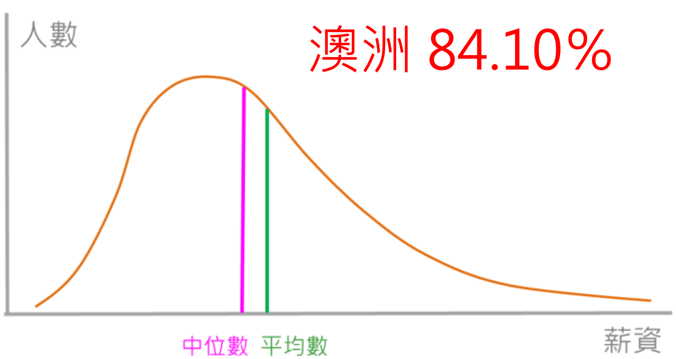
不同的母體分配（延續假設案例）

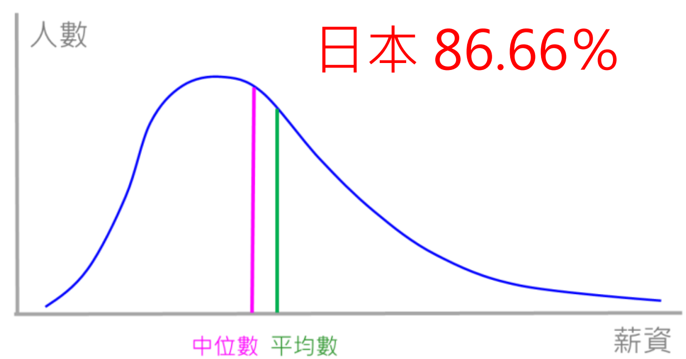


美國 74.90%



偏度與峰度都會有差異，不一定可以在同一個座標軸呈現





母體分配、樣本分配與抽樣分配有何不同？

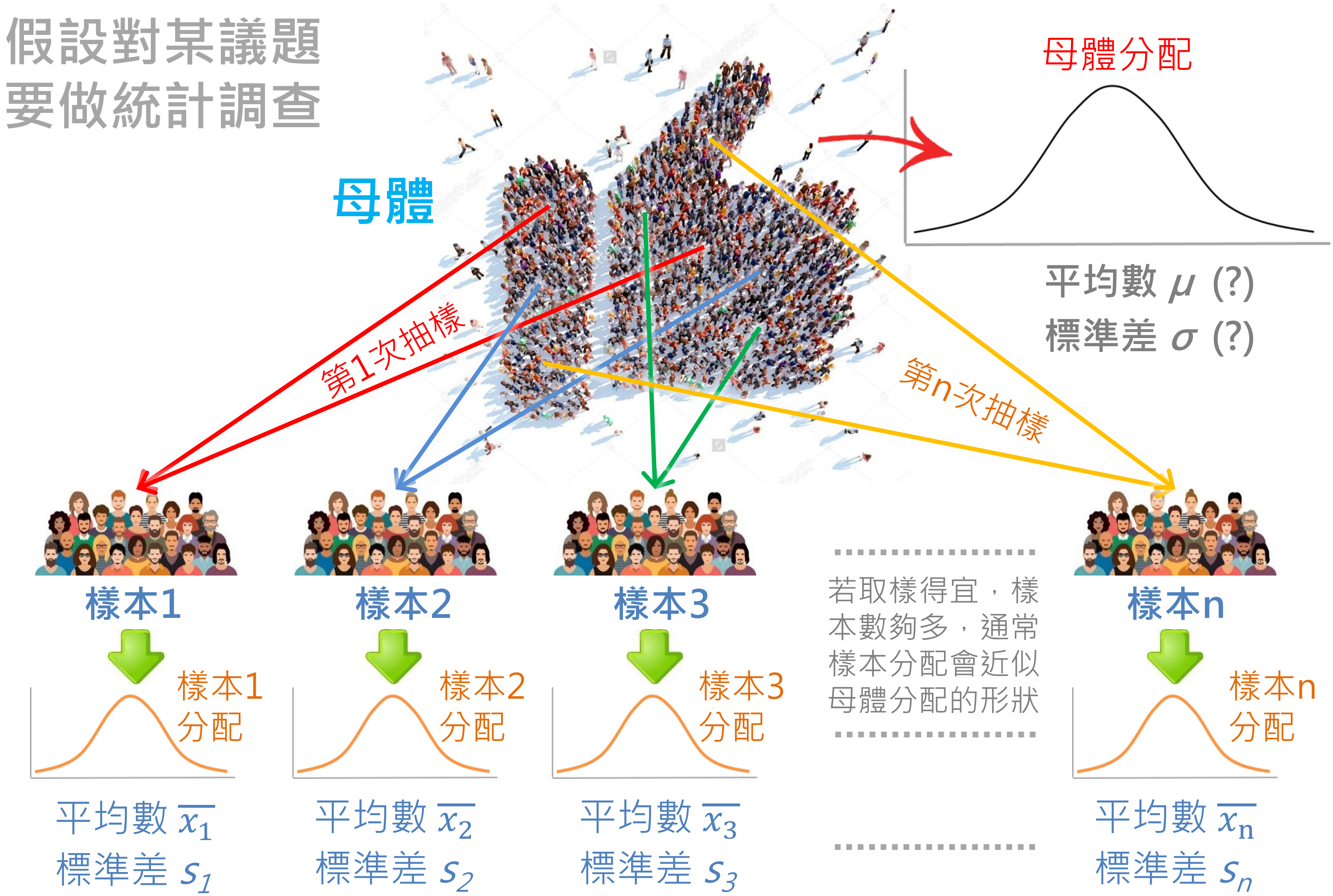
相同的定義

（對「分配」的定義都相同）

不同的含意

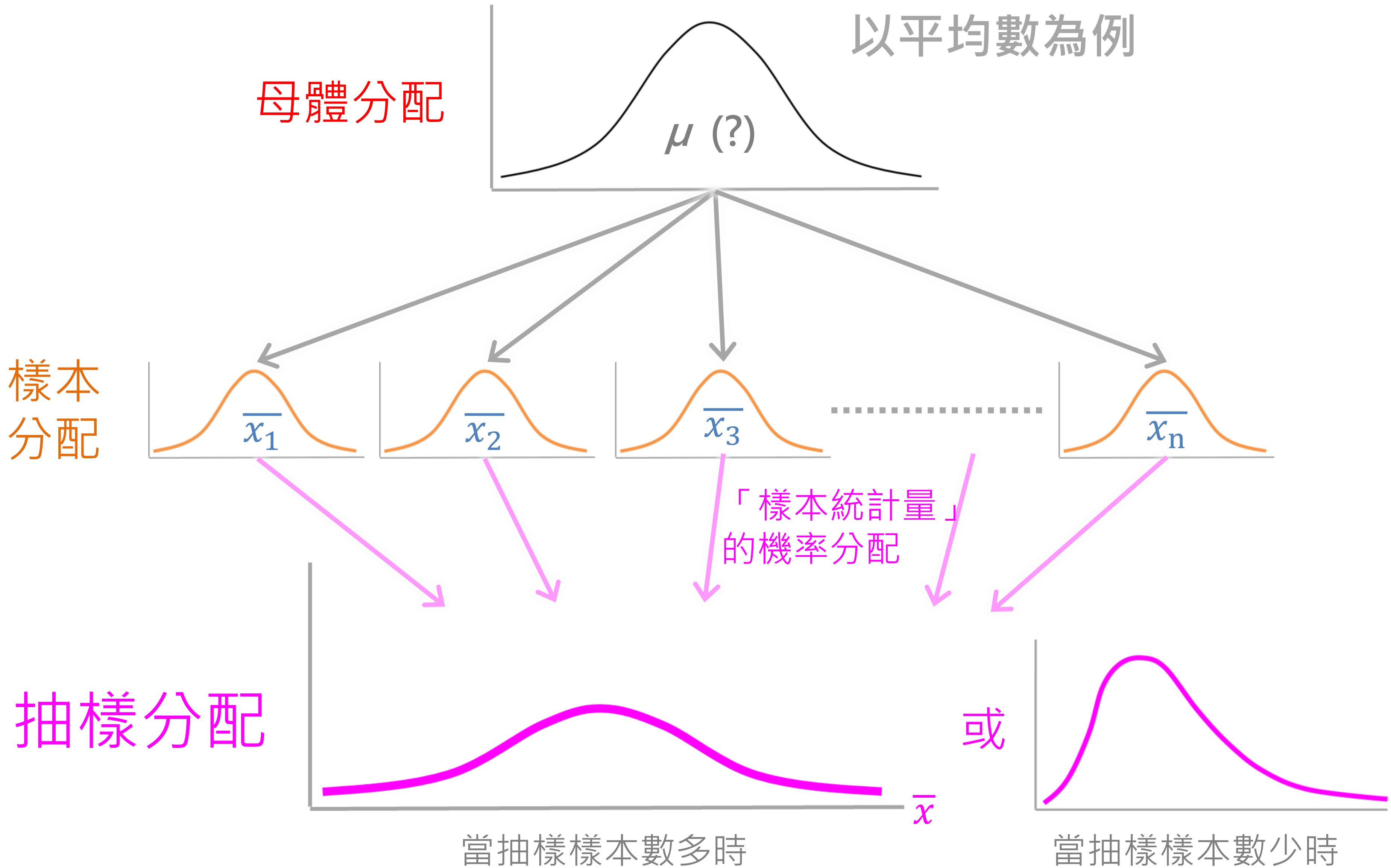
（母體、樣本、抽樣的含意不同）

母體分配與樣本分配



假設對某議題
要做統計調查
母體
母體分配
平均數 μ (?)
標準差 σ (?)
第1次抽樣
第n次抽樣
樣本1
樣本2
樣本3
樣本n
樣本1分配
樣本2分配
樣本3分配
樣本n分配
若取樣得宜，樣本數夠多，通常樣本分配會近似母體分配的形狀
平均數 $\overline{x_1}$
標準差 s_1
平均數 $\overline{x_2}$
標準差 s_2
平均數 $\overline{x_3}$
標準差 s_3
平均數 $\overline{x_n}$
標準差 s_n

樣本分配與抽樣分配


以平均數為例
母體分配
μ (?)
樣本分配
$\overline{x_1}$
$\overline{x_2}$
$\overline{x_3}$
$\overline{x_n}$
「樣本統計量」
的機率分配
抽樣分配
$\overline{x}$
或
當抽樣樣本數多時
當抽樣樣本數少時

統計主要使用的分配

統計對象	型態	分配類型	
母體分配 樣本分配	離散型 機率分配	均勻分配	二項分配
		多項分配	負二項分配
		超幾何分配	幾何分配
		伯努利分配	波松分配
	連續型 機率分配	連續均勻分配	常態分配
		標準常態分配	指數分配
		Gamma分配	卡方分配
抽樣分配	機率分配	樣本比例差（$\lvert\hat{p}-p\rvert$）	
		平均數～	
		Z分配	t分配
		變異數～	
		單母體變異數	卡方分配
		兩母體變異數比	F分配

The End