

學習分配的目的性

學「分配」要做什麼？



番外篇

先用比喻來說明

各國種族文化大不同



送禮習慣

新年習俗

打招呼

家庭相處

飲食口味

生活作息

祭拜儀式

價值觀

假設對不同族群的異性互有好感

你可能想要知道.....

他可能是那個族群的？

有沒有那個族群的特質？

跟自己的差異有多大？

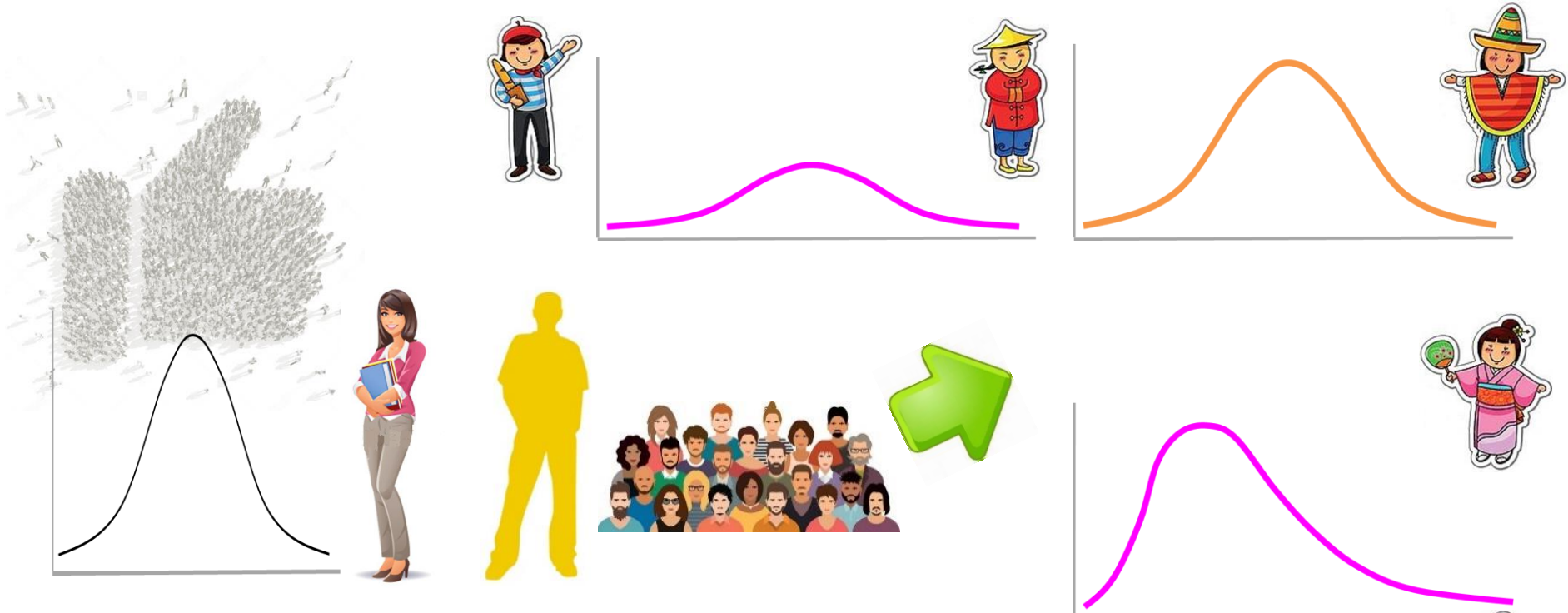
有沒有可能在一起？



心儀的對象



「族群」與「分配」的比喻



他可能是那個族群的？（哪種分配樣式？）

有沒有那個族群的特質？（平均數？變異數？偏度？峰度？）

跟自己的差異有多大？（在某個範圍內的相似或相異的機率？→估計）

有沒有可能在一起？（透過樣本訊息與機率原理做出接受或拒絕的統計方法 → 假設檢定）

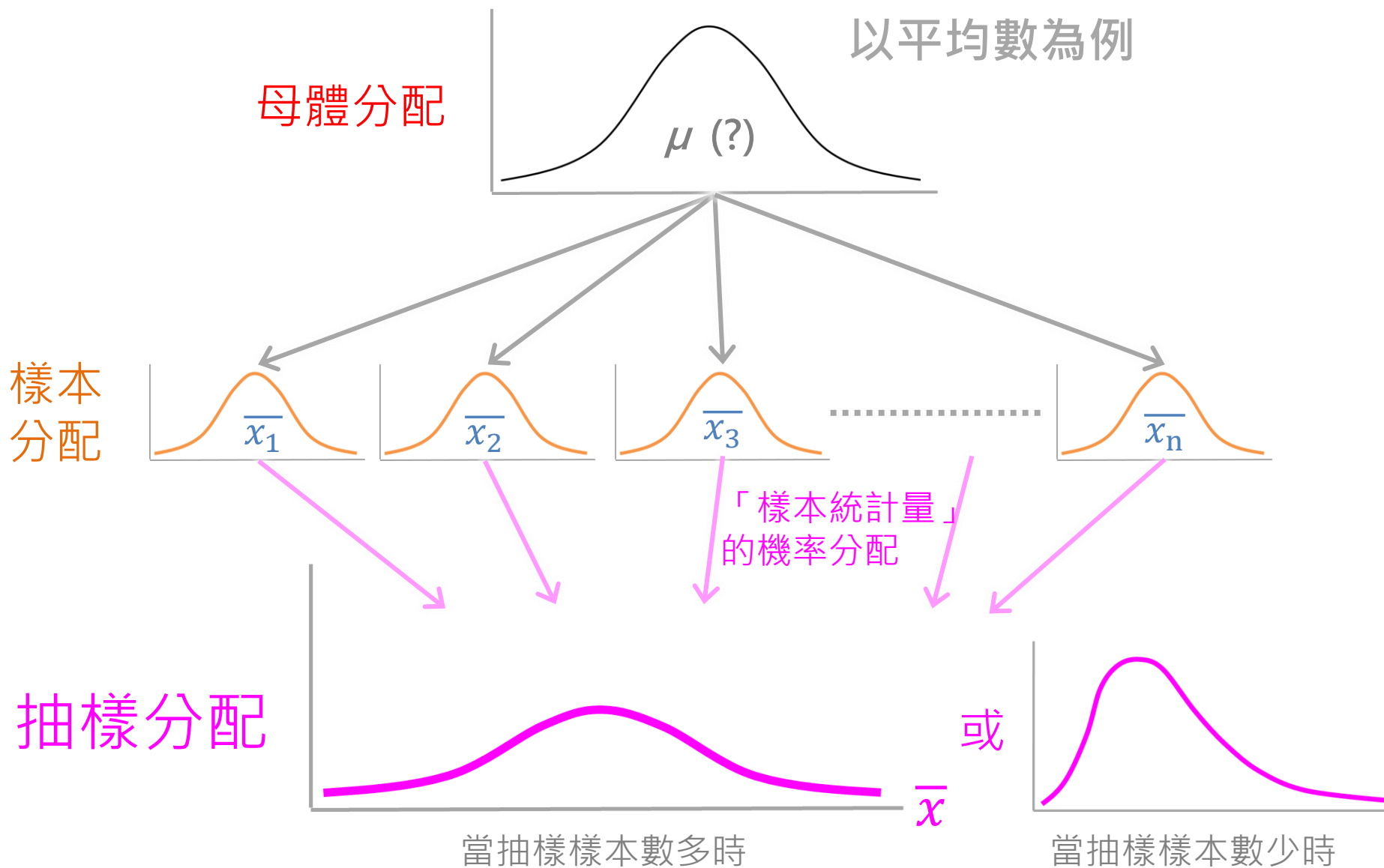




回到統計

學習分配的目的性.....

母體分配、樣本分配與抽樣分配



我們對怎樣的議題會有興趣？

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 多項分配 超幾何分配 伯努利分配	二項分配 負二項分配 幾何分配 波松分配
	連續型 機率分配	連續均勻分配 標準常態分配 Gamma分配	常態分配 指數分配 卡方分配
抽樣分配	機率分配	樣本比例差 ($ \hat{p} - p $) 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	

期望值 (μ , expected value)

數學符號 μ (mu)

Expected Value for a discrete random variable is:

$$E(X) = \sum x_i \cdot p(x_i) \quad (\text{離散型})$$

Expected Value for a continuous random variable is:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (\text{連續型})$$

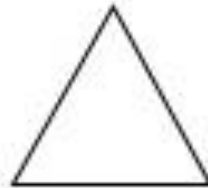
The mean describes where the probability distribution is centered.

實際上，期望值就是算數平均數

期望值案例1







- Recall the following probability distribution of ER arrivals:







x	10	11	12	13	14
$P(x)$.4	.2	.2	.1	.1



$$\sum_{i=1}^5 x_i p(x) = 10(.4) + 11(.2) + 12(.2) + 13(.1) + 14(.1) = 11.3$$

期望值案例2

\$1 Payout	Your Roll						
	Probability	1:6	1:6	1:6	1:6	1:6	1:6
	Payout	-\$1	-\$1	-\$1	-\$1	-\$1	+\$1
	$E(R) = \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (+1)$ $E(R) = \left(\frac{5}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) (+1) = -\frac{2}{3} = -\$0.67 \text{ per bet} = -67\% = \text{Bad Bet!}$						

\$10 Payout	Your Roll						
	Probability	1:6	1:6	1:6	1:6	1:6	1:6
	Payout	-\$1	-\$1	-\$1	-\$1	-\$1	+\$10
	$E(R) = \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) \cdot (+10)$ $E(R) = \left(\frac{5}{6}\right) \cdot (-1) + \left(\frac{1}{6}\right) (+10) = +\frac{5}{6} = +\$0.83 \text{ per bet} = +83\% = \text{Good Bet!}$						

期望值的線性關係

設 $Y = a + bX$

$$E[Y] = \sum_{x \in \mathcal{R}_X} (a + bx)p_X(x) \quad (\text{using the transformation theorem})$$

(離散型)

$$= \sum_{x \in \mathcal{R}_X} ap_X(x) + \sum_{x \in \mathcal{R}_X} bxp_X(x)$$

$$= a \sum_{x \in \mathcal{R}_X} p_X(x) + b \sum_{x \in \mathcal{R}_X} xp_X(x)$$

$$= a + b \sum_{x \in \mathcal{R}_X} xp_X(x) \quad (\text{because probabilities sum up to 1})$$

$$= a + bE[X] \quad (\text{by the definition of } E[X])$$

$$E[Y] = \int_{-\infty}^{\infty} (a + bx)f_X(x)dx \quad (\text{using the transformation theorem})$$

(連續型)

$$= \int_{-\infty}^{\infty} af_X(x)dx + \int_{-\infty}^{\infty} bxf_X(x)dx$$

$$= a \int_{-\infty}^{\infty} f_X(x)dx + b \int_{-\infty}^{\infty} xf_X(x)dx$$

$$= a + b \int_{-\infty}^{\infty} xf_X(x)dx \quad (\text{because probability densities integrate to 1})$$

$$= a + bE[X] \quad (\text{by the definition of } E[X])$$

變異數(σ^2 , variance)

(離散型)

數學符號 σ (sigma)

$$\text{Var}(X) = \text{E}[(X - \mu)^2]$$

$$\begin{aligned}\text{Var}(X) &= \text{E}[X^2 - 2X\text{E}[X] + (\text{E}[X])^2] \\ &= \text{E}[X^2] - 2\text{E}[X]\text{E}[X] + (\text{E}[X])^2 \\ &= \text{E}[X^2] - (\text{E}[X])^2\end{aligned}$$

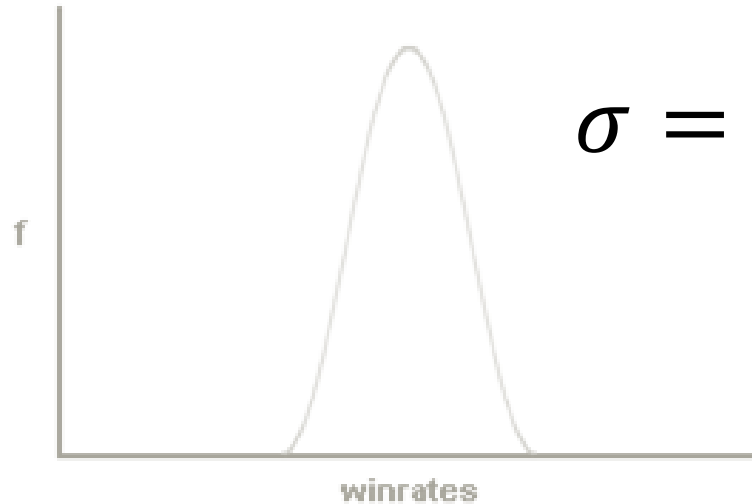
(連續型)

$$\text{Var}(X) = \sigma^2 = \int (x - \mu)^2 f(x) dx = \int x^2 f(x) dx - \mu^2$$

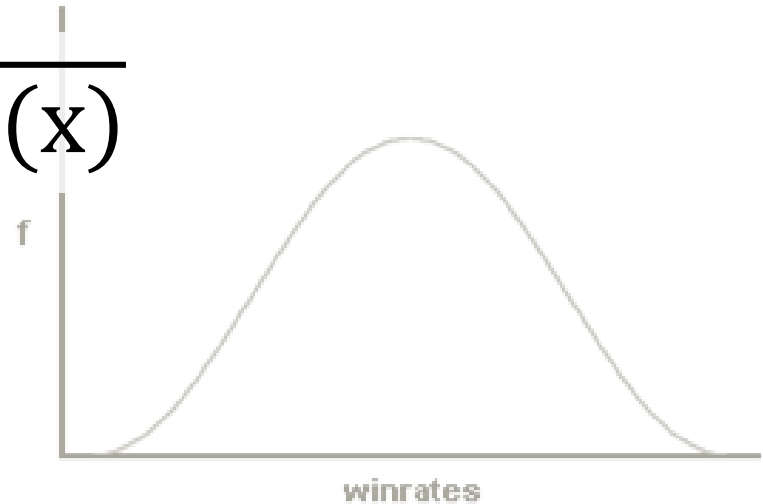
一個隨機變量的變異數，描述的是它的離散程度，也就是該變量離其期望值的距離。

標準差(σ , standard deviation)

Low Standard Deviation



High Standard Deviation



$$\sigma = \sqrt{V(x)}$$

- 標準差定義為變異數的算術平方根。
- 標準差是一組數值自平均值分散開來的程度的一種離散程度測量觀念。
- 一個較大的標準差，代表大部分的數值和其平均值之間差異較大；一個較小的標準差，代表這些數值較接近平均值。

母體的標準差

$$\begin{aligned}\sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2} & \sum_{i=1}^N (X_i - \mu)^2 &= \sum_{i=1}^N (X_i^2 - 2X_i\mu + \mu^2) \\ & & &= \left(\sum_{i=1}^N X_i^2 \right) - \left(2\mu \sum_{i=1}^N X_i \right) + N\mu^2 \\ &= \sqrt{\frac{1}{N} \left(\sum_{i=1}^N X_i^2 \right) - \frac{1}{N} N\mu^2} & &= \left(\sum_{i=1}^N X_i^2 \right) - 2\mu(N\mu) + N\mu^2 \\ & & &= \left(\sum_{i=1}^N X_i^2 \right) - 2N\mu^2 + N\mu^2 \\ &= \sqrt{\frac{\sum_{i=1}^N X_i^2}{N} - \mu^2} & &= \left(\sum_{i=1}^N X_i^2 \right) - N\mu^2\end{aligned}$$

根號裡面的簡易口訣為：

「平方和的平均」減去「平均的平方」。



The End