單元 10：基本統計學的量度及其釋義

10.1 集中趨勢的量度：平均數、加權平均數、中位數和眾數

本部分基本上是溫習從前學過的課題。教師應與學生重溫各種數據的圖示法，以及各種集中趨勢的量度方法。爲要引起學生興趣，教師應多給與一些實際生活例子。

日常生活例子：(摘錄自《生活與統計》，由政府統計處製作)

(i) 一班學生的平均高度是 1.73 米。
(ii) 1994 年每天的平均最高氣溫是 25.9 °C。
(iii) 1994 年第三季家庭收入的中位數是 $15,000。
(iv) 1993 年男性結婚年齡的中位數是 30 歲 (只計算首次結婚的年歲)
(v) 1991 年人口普查的資料顯示，家庭人數的泉數值是 4 人。
(vi) 貨品生產商需知道每件產品的眾數值。如恤衫製造商會製造較多是數尺碼的恤衫。

以下是另外一些例子：

例一
試利用每天在報章上刊登的數據 (如：股價、世界各城市氣溫等)，比較其集中趨勢 (平均數、中位數和眾數)。

例二
記錄每位學生每星期所得的零用錢數目，數額以 $10 爲單位。試找出零用錢的平均數、中位數和眾數。哪一個數據較能代表學生每月所得的零用錢 ? 學生可從這個例子察覺數據有時會有多於一個眾數。

例三
一公司員工的每年收入如下：

董事	$ 2,400,000	經理	$ 2,000,000
營業員	$ 300,000	總服務員	$ 150,000
3 名服務員	每位 $ 100,000	2 名秘書	每位 $ 90,000
1 名見習員	$ 80,000		

求這 10 名員工薪金的眾數、中位數及平均數。哪個數值最能代表員工的平均薪金 ? 試解釋答案。

加權平均數

日常生活常會用到加權平均數的概念。分組數據的平均值實際上是將各組中值乘以各組頻數後得出的平均數。考試分數的計算亦運用了加權平均數的概念。

加權平均數廣泛應用在各種指數的計算上，如物價指數等。

指數

指數是一個用來表示某些數據相對於某時期的變化和走勢的數字。作爲比較的基準時期叫做基期，其指數通常定爲 100。指數是以基期的百分數表示。

香港有各種各樣的指數，例如：
(i) 甲類消費物價指數
(ii) 類消費物價指數
(iii) 恒生物價指數

各採用不同的加權數值去反映不同家庭的不同消費模式。

例一
甲類消費物價指數覆蓋了全港市區家庭的 50%，這些家庭在 1989 年 9 月至 1990 年 9 月期間的每月支出是在 $ 2,500 至 $ 9,999 之間。爲了顯示每一時期各種食物及服務的一般水準，我們對每個時期用一個百分數 (即指數) 表示，以觀察其總體上的變化。

例二
恒生指數在 1964 年 7 月 31 日爲 100，在 1997 年 9 月 29 日，已上升至 14,864，顯示總市值有超過 100 倍的增長。

物價指數

商品	基期價格	時期 t 的價格	權數
1	c_{01}	c_{t1}	w_1
2	c_{02}	c_{t2}	w_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	c_{0k}	c_{tk}	w_k

t 期物價加權平均指數

$$= \frac{\text{基期價格的加權平均數}}{\text{基期價格的加權平均數}} \times 100$$

$$= \frac{\dfrac{w_1c_{t1} + w_2c_{t2} + w_3c_{t3} + \cdots + w_kc_{tk}}{w_1 + w_2 + w_3 + \cdots + w_k}}{\dfrac{w_1c_{01} + w_2c_{02} + w_3c_{01}3 + \cdots + w_kc_{0k}}{w_1 + w_2 + w_3 + \cdots + w_k}} \times 100$$

$$=\frac{w_1c_{t1}+w_2c_{t2}+w_3c_{t3}+\cdots+w_kc_{tk}}{w_1c_{01}+w_2c_{02}+w_3c_{03}+\cdots+w_kc_{0k}}\times 100$$

例三

假設以 1992 年爲基期，下表列出了 1992 年和 1995 年幾種主要食物及零食的價格，求 1995 年的物價指數。

商品	1992年的價格	1995年的價格	權數
米	28	30	5
豬肉	21	26	4
魚	30	36	4
蔬菜	6	8	3
零食	4	6	2

物價指數

$$=\frac{5\times 30+4\times 26+4\times 36+3\times 8+2\times 6}{5\times 28+4\times 21+4\times 30+3\times 6+2\times 4}\times 100$$

= 117.3

建議指數研究課題：

(i) 個人的消閑物品消費指數。
(ii) 學生交通消費指數。

集中趨勢量度的應用

我們用以下例子說明：

陳先生以更換破爛窗子爲業。根據過往數月的工作記錄，他的女兒替他計算出更換一個窗子所需時間的結果如下：

平均數 = 2.0 小時
中位數 = 1. 8 小時
眾數 = 1. 5 小時

這些數字對陳先生有甚麼作用呢 ?

(i) 颱風過後，一所學校的 100 隻窗給擊碎了。陳先生大約需要 $100 \times 2.0 = 200$ 小時去更換那些窗子。這情況所用的時間採用平均數來計算。

(ii) 若問陳先生一般需要多少時間去更換一個窗子，他通常會利用眾數作爲答案。

(iii) 陳先生日常更換窗戶是以件工計算。現在公司要求他在 2 小時內更換一個窗子，陳先生應否接納這份工作 ? 爲甚麼 ?

教師可利用其他例子帶出平均數、中位數及眾數的優點及缺點。

10.2 離差的量度

教師可提出兩組平均數及中位數相同但分佈狀態不一樣的數據，讓學生看出，單以集中趨勢的量度去描述數據很明顯並不足夠。如何量度散佈在平均數四周的數據呢 ? 這種量度的方法稱爲離差的量度。

離差量度的種類

以下是常用的離差量度：

(i) 分佈域

(ii) 四分位數間距或半四分位數間距

(iii) 方差和標準差

以上各項的定義可用生活化實例及圖像說明。跟很多數學概念一樣，用圖像顯示統計數據比用符號或數字較易令學生明白和掌握。

(i) 分佈域

日常生活例子：天氣報告；氣溫相差；股市波動；潮水漲退等。

兩組有相同的分佈域但分佈不同的數據，可以用圖像來顯示。

(ii) 四分位數間距或半四分位數間距

日常生活例子：薪金；考生入學試的成績等。

我們可根據數據的性質，用插值法來估計四分位數之值。

註：除可運用四分位數，也可採用百分位數來計算。

(iii) 方差及標準差

以上提及的分佈域、四分位數間距、半四分位數間距，並沒有盡用數據內的所有數值。當某些數據的數值改變時，離差值並沒有因應改變。用來描述所有數據分佈狀態的一個重要計算方法是方差，或稱爲標準差。

<u>例一</u>

找出不同年齡的男孩及女孩高度的標準差，並作出適當的結論。

<u>例二</u>

以下是兩位學生做實驗時所量度得到的兩組數據：

	酸的體積 (cm^3)									
學生 A	8	12	7	9	3	10	12	11	12	14
學生 B	7	6	7	15	12	11	9	9	13	11

(i) 求每一學生所量度酸的體積之平均值。

(ii) 求酸的體積之標準差。

(iii) 哪組數據較可靠 ?

比較不同離差量度

教師應詳細說明每種離差量度的利弊。

10.3 頻數分佈、累積頻數分佈及其圖像顯示和釋義

教師可用生活實例去複習這課題。

分組數據和不分組數據的異同，宜作重點說明。每種統計圖均有其特色，教師可透過各種不同的圖像作示範，從而說明某類圖像是否適用於表達某些資料與結論。重點應放在理解圖表和圖表在預測方面的功用。

可利用電腦軟件將已知數據繪畫成不同圖像。當學生已經掌握用各種圖像表達數據的方式後，可嘗試利用電腦分析各種形式的異同，以及優劣的地方。教師可運用電腦作爲教學工具，使學生便於明白。

以下是一些課題建議，作爲學生研究的項目：

1. 題目：人體體溫
   問題：是否所有人皆有相同體溫 ? 男性與女性的體溫是否相約 ? 是否女性的體溫較男性高 ? 女性的體溫波幅是否較大 ?
   數據搜集：用人體溫度計量度 50 – 100 人的體溫。大部份溫度計以攝氏 (°C) 單位表示，量度溫度須準確至 0.1 °C。

2. 題目：路口交通情況
   目的：研究某繁忙路口於一分鐘內汽車通過的數目。
   數據搜集：選擇一交通流量大的路口。決定以何種車輛作爲統計數據及統計的地點所在。數算該類車輛於一分鐘內在同一方向通過該路口的數目。統計時間爲一小時。若時間容許，可選擇一交通流量受阻的地方，例如在紅綠燈附近，或十字路口。重複以上步驟並作記錄。 (若作爲學生的研究項目，可選擇不同地點給不同組別的學生。)

3. 題目：中六生的零用錢數目
   問題：中六男生或女生每星期所得的零用錢是否相約 ?
   數據搜集：問 50 位中六男生及女生每星期的零用錢數目。若他們沒有零用錢，記錄那些答案爲零。分別記錄女生和男生的零用錢的數目。

4. 設計及搜集關於一些課題的意見 (例如：流行歌曲、政治團體、教育學歷等。)

5. 究竟在某一快餐店購買 “中薯條”，會有多少條薯條呢 ? 學生搜集這些資料時，會發覺需要處理碎薯條的情況。很快地，班內會有不同的處理方法。我們可以怎樣描述那些答案呢 ? 你發現甚麼情況呢 ? 有經驗的學生可能會討論薯條之長度或形狀分佈。薯條的形狀大小，要怎樣才不會被人投訴 ? 最好購買那一類包裝的薯條：大、中或小 ? 這快餐店與其他店舖的比較又如何呢 ?

10.4 箱形圖

箱形圖 (亦稱框綫圖) 可用作展示一組數據的特徵，包括最小值、下四分位數、中位數、上四分位數、最大值。

在水平圖尺上繪製箱形圖的步驟如下：

步驟 1. 在圖尺上繪畫一個窄長的方箱，由下四分位數伸延至上四分位數；
步驟 2. 在方箱內中位數的位置上畫一垂直綫段；
步驟 3. 從方箱的左邊畫一綫段到最小值；
步驟 4. 從方箱的右邊畫一綫段到最大值。

步驟 3 與步驟 4 所畫的兩綫段稱爲框綫。

圖 10.4(i) 顯示一個箱形圖的大致形狀。

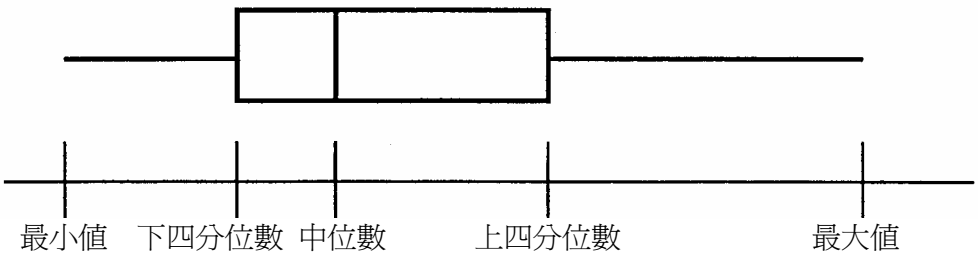


圖 10.4 (i)

用者可隨意選定箱形圖的方箱闊度，絲毫不會影響該圖所表達的意思。

下面例子說明箱形圖的製作。

<u>例一</u>
以下是 20 位足球員每星期用在購買汽水的支出：

20 位足球員用在
購買汽水的支出 (元)

26	34	31	36	16
27	32	21	43	41
30	6	15	38	27
58	35	28	21	20

從這些數字得知
最大值 = 58
上四分位數 = 35.5
中位數 = 29
下四分位數 = 21
最小值 = 6

繪成箱形圖如下 (圖 10.4(ii))：

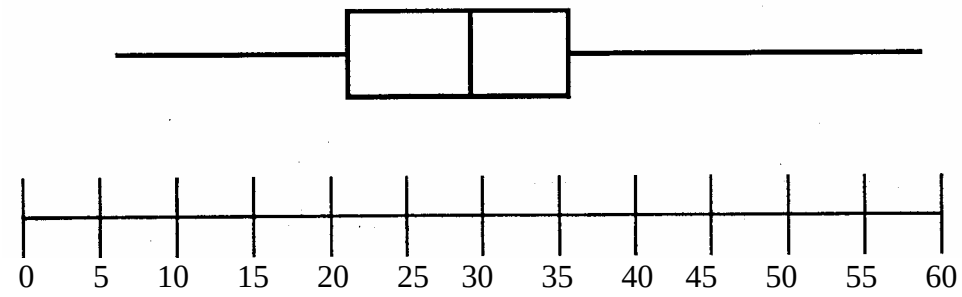


圖 10.4(ii) 購買汽水費用箱形圖

把上述製圖的步驟作出相應的調整後，箱形圖也可以在垂直圖尺上繪製 (參照圖 10.4(i))

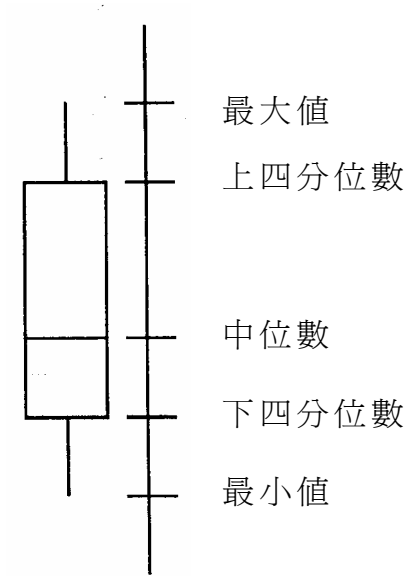


圖 10.4(iii)

箱形圖可以表示一組數據的中心位置、散佈情況及對稱性。

1. 中位數上的綫段表示數據的中心位置。
2. 方箱的長度顯示中間一半數據的散佈情況。
3. 兩綫段的長度顯示數據的第一個及第四個四分之一部分的散佈情況。
4. 從圖的形狀即可看出數據分佈的對稱程度。
   如果圖是對稱的，有關數據很可能是對稱的。偏離對稱的圖可反映出有關的數據也有類似的偏離 (見圖 10.4(iv) 及 10.4(v))。

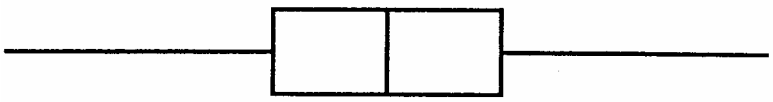


圖 10.4(iv) 對稱於中位數的數據分佈

圖 10.4(v) 非對稱於中位數的數據分佈

箱形圖用來比較多組數據，也很有效。可用作比較不同數據組的中心位置、散佈情況及對稱性。

圖 10.4(vi) 的箱形圖展示三個城市的工資分佈。

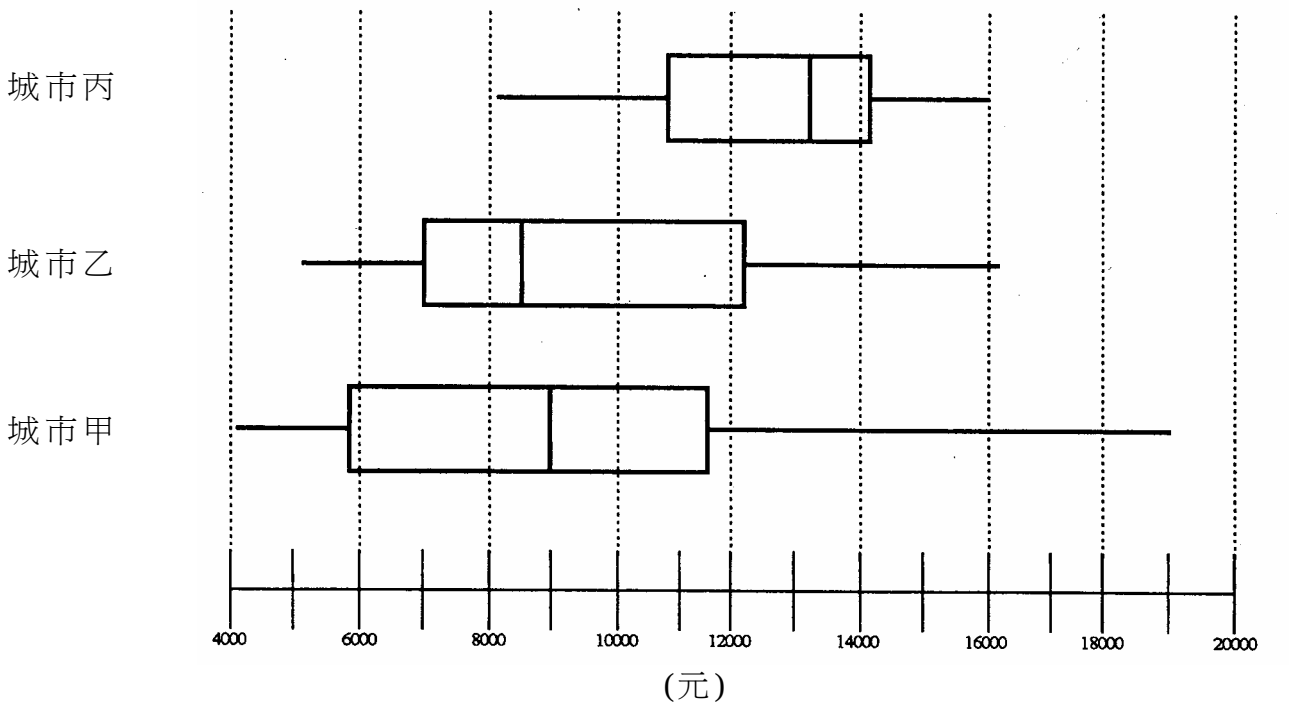


圖 10.4(vi) 三個城市的工資分佈

由圖 10.4(vi) 可知，城市丙的工資中位數最高，城市甲的工資離差最大，城市乙近半數在職人士的工資低於城市丙的最低工資。

10.5 幹葉圖

幹葉圖是一種結合圖與表以表達數據的方法，以數據來替代直方圖中的直方形，基本上是一已作改良的分組頻數分佈。我們採用數據的最後單位(稱爲「葉」) 來代替畫記，用來顯示觀察所得的數值。這些觀察值是根據「葉」之前的數字 (稱爲「幹」) 作排序，由小至大依次排列在葉的左方。

以下是利用 40 位學生的數學測驗成績，來顯示幹葉圖的製作：

49	73	58	59	40	61	45	61
51	43	80	55	81	52	54	45
58	53	55	38	59	72	46	68
42	65	65	64	29	48	39	67
47	56	66	48	42	56	51	53

步驟 1. 畫一垂直綫。

步驟 2. 選擇數據的前置數字爲幹。

步驟 3. 將幹從小至大排列，並列於該垂直綫的左方。

步驟 4. 檢視這些數據，將最後的數字 (葉) 由小至大排列在對應幹的右方。

學生應能掌握描繪及解釋幹葉圖的方法。

幹 (十位值)	葉 (個位值)
2	9
3	89
4	02235567889
5	1123345566889
	9
6	11455678
7	23
8	01

圖 10.5 (i)

例一

幹 (百位值)	葉 (十位值)
1	11 23 47
	48
2	23 27
3	55 63 80
4	21 23
5	00

圖 10.5(ii)

在上圖，幹是百位數及葉是十位以下數字。

幹葉圖所顯示的數據爲：

111, 123, 147, 148, 223, 227, 355, 363, 380, 421, 423, 500

例二

以下爲某家庭於冬天的九個星期裏每星期的耗電量 (單位爲 KWh)：
338，354，341，353，351，341，353，346，341。

根據資料，耗電量的百位數字每個都是 3。再觀察數據的十位值，數值爲 3、4 或 5 。所以，我們選取數字 33、34 及 35 爲幹，放在垂直綫的左方，個位值爲葉，放在垂直綫的右方。最後把數據依遞增次序排列，製成幹葉圖，如圖 10.5(iii) 所示。

幹 (十位值)	葉 (個位值)
33	8
34	1 1 1 6
35	1 3 3 4

圖 10.5 (iii)

例三

11 位學生的中、英文科考試成績如下：

英文科： 23， 39， 40， 45， 51， 55， 61， 64， 65， 72， 78
中文科： 37， 41， 44， 48， 58， 61， 63， 69， 75， 83， 89

可繪製並列幹葉圖來比較兩科成績的表現。

英文科

幹 (十位值)	葉 (個位值)
2	3
3	9
4	0 5
5	1 5
6	1 4 5
7	2 8
8	

中文科

幹 (十位值)	葉 (個位值)
2	
3	7
4	1 4 8
5	8
6	1 3 9
7	5
8	3 9

圖 10.5(iv)

利用背靠背的幹葉圖可以更清楚反映兩者的比較。

英文科 葉 (個位值)	幹 (十位值)	中文科 葉 (個位值)
3	2	
9	3	7
5 0	4	1 4 8
5 1	5	8
5 4 1	6	1 3 9
8 2	7	5
	8	3 9

圖 10.5(v)

對於某些數值特別小或特別大的數據，爲要令到幹葉圍看來較美觀，可以另外寫出這些數據。用以下的數據爲例子，如圖 l 0.5(vi) 或圖 10.5(vii) 所示。

12.0	21.2	23.1	24.8
22.0	25.9	29.8	31.9
21.2	21.6	22.8	22.5
26.3	25.1	21.7	21.6
23.5	27.0	38.5	28.5

幹 (個位值)	葉 (十分位值)
LO = 12.0	
21	2 2 6 6 7
22	0 5 8
23	1 5
24	8
25	1 9
26	3
27	0
28	5
29	8
HI = 31.9, 38.5	

圖 10.5(vi)

幹 (個位值)	葉 (十分位值)
LO = 12.0	
21	2 2 6 6 7
22	0 5 8
23	1 5
24	8
25	1 9
26	3
27	0
28	5
29	8
30	
31	9
HI = 38.5	

圖 10.5(vii)

10.6 總體參數和樣本統計量的初步認識

本節介紹**總體**、**總體參數**，**樣本**及**樣本統計量**的概念。從總體抽取一些樣本，這些樣本可以適當地代表總體，從而取得估計量的總體參數及作出推論。可用例子介紹總體、樣本、樣本方差及樣本平均數這幾個名詞。

總體及樣本

例一

年齡分佈：　　某中學學生的年齡分佈。
總體：　　該校的所有學生。
樣本：　　隨機在該校抽取 100 名學生。

例二

電撞見觀眾：進行一項調查於下午六時至晚上十時最多人觀看的電視節目。
總體：　　下午六時至晚上十時期間，所有電視觀眾。
樣本：　　從總體電視觀眾中隨機抽取 800 人作爲訪問對象。

例三

晶質控制：工廠所生產的電燈泡壽命 o
總體：　　全部電燈泡產品。
樣本：　　隨機抽取 30 個電燈泡加以測試。

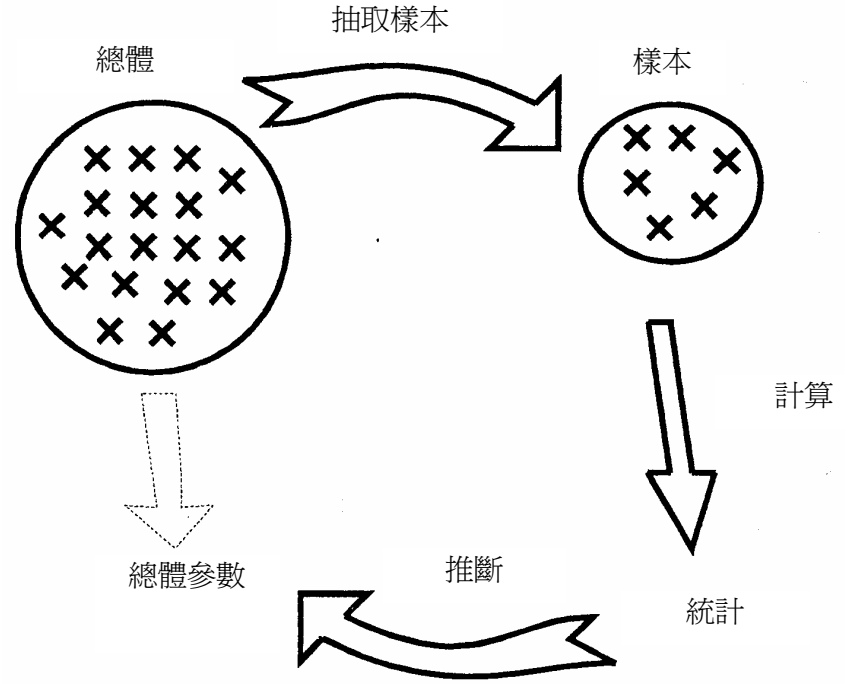
例四

某生產錄影機公司的客戶服務部經理有意作一項調查，看看在過去 12 個月內購買錄影機的顧客對產品是否滿意。他計劃根據 1,425 名顧客所交回的保養證去進行調查。

i)　試描述經理所想要了解的調查總體及樣本。
ii)　試描述該經理有興趣搜集的數據。

總體參數是描述總體的特色，而樣本統計量則是由總體抽取樣本得出特色的一些計算值。

教師可以利用下圖說明各項的關係：



樣本平均數及樣本方差

學生應知道：

(i) 樣本平均數 $\bar{x}=\frac{1}{n}\sum_{i=1}^{k}f_i x_i$， 其中 $n=\sum_{i=1}^{k}f_i$

(ii) 樣本方差 $s^2=\frac{1}{n-1}\sum_{i=1}^{k}f_i\left(x_i-\bar{x}\right)^2$，其中 $n=\sum_{i=1}^{k}f_i$

(iii) 當樣本數目增至無限大時，樣本平均數 $\bar{x}$ 及樣本方差 s^2 分別趨向總體平均數 μ 及總體方差 σ^2。

(iv) 對於有限的總體來說，總體方差

$\sigma^2=\frac{1}{N}\sum_{i=1}^{k}f_i\left(x_i-\mu\right)^2$，其中 N 是總體數目。

學生並不需要明白**樣本平均數的平均值**及**樣本平均數的方差**的意義。

<u>例一</u>

一硬幣的兩面各刻上數字 3 及 4 。這硬幣是有偏差的，平均來說，每 5 次投擲中有 4 次出現數字 3 及 1 次出現數字 4 。因此可假設總體只有這幾個數：3，3，3，3，4。

總體平均數 $\mu=\frac{3+3+3+3+4}{5}=3.2$

總體方差 $\sigma^2=\frac{4\times(3-3.2)^2+(4-3.2)^2}{5}=0.16$

若學生不知道這硬幣有偏差，他們投擲這硬幣 3 次，所有可能出現的樣本如下：(3, 3, 3), (3, 3, 4), (3, 4, 3), (4, 3, 3), (3, 4, 4), (4, 3, 4), (4, 4, 3), (4, 4, 4).

以樣本 (3, 3, 3) 爲例：

樣本平均數 $\bar{x}=\frac{3+3+3}{3}=3$

樣本方差 $s^2=\frac{(3-3)^2+(3-3)^2+(3-3)^2}{3-1}=0$

以另一樣本 (3, 4, 4) 爲例：

樣本平均數 $\bar{x}=\frac{3+4+4}{3}=3\frac{2}{3}$

樣本方差 $s^2=\frac{(3-3\frac{2}{3})^2+(4-3\frac{2}{3})^2+(4-3\frac{2}{3})^2}{3-1}=\frac{1}{3}$

教師憑此可指出：樣本平均數和樣本方差與總體平均數和總體方差並不一定相同。