

# 統計學（一）

## 第一章 (Introduction)

授課教師：唐麗英教授

國立交通大學 工業工程與管理學系

聯絡電話：(03)5731896

e-mail：[litong@cc.nctu.edu.tw](mailto:litong@cc.nctu.edu.tw)

2013

☆ 本講義未經同意請勿自行翻印 ☆

# 本課程內容參考書目

- 教科書

- P. Newbold, W. L. Carlson and B. Thorne(2007). *Statistics for Business and the Economics*, 7<sup>th</sup> Edition, Pearson.

- 參考書目

- Berenson, M. L., Levine, D. M., and Krehbiel, T. C. (2009). *Basic business statistics: Concepts and applications*, 11<sup>th</sup> Edition Prentice Hall.
- Larson, H. J. (1982). *Introduction to probability theory and statistical inference*, 3<sup>rd</sup> Edition, New York: Wiley.
- Miller, I., Freund, J. E., and Johnson, R. A. (2000). *Miller and Freund's Probability and statistics for engineers*, 6<sup>th</sup> Edition, Prentice Hall.
- Montgomery, D. C., and Runger, G. C. (2011). *Applied statistics and probability for engineers*, 5<sup>th</sup> Edition, Wiley.
- Watson, C. J. (1997). *Statistics for management and economics*, 5th Edition. Prentice Hall.
- 唐麗英、王春和（2013），「從範例學MINITAB統計分析與應用」，博碩文化公司。
- 唐麗英、王春和（2008），「SPSS 統計分析」，儒林圖書公司。
- 唐麗英、王春和（2007），「Excel 統計分析」，第二版，儒林圖書公司。
- 唐麗英、王春和（2005），「STATISTICA與基礎統計分析」，儒林圖書公司。

- **Statistics (統計學)**

- 統計學為蒐集、整理、展示、分析、解釋資料，並由樣本推論群體，使在不確定的情況下做成決策的科學方法

- **population (群體)**

- 由具有共同特性之個體所組成的整體

- **sample (樣本)**

- 群體之一部分

- **parameter (參數)**
  - 由**群體**資料所計算之**群體表徵值**
- **statistic (統計量)**
  - 由**樣本**資料所計算之**樣本表徵值**
- **The objective of Statistics (統計學的目的)**
  - 由**樣本**資訊推論母體參數

- 例 1：

某製程工程師欲由100片隨機抽出之晶圓來估計晶圓之厚度。請指出此例欲研究之群體、樣本、參數及統計量各為何。

- **統計學的範圍**

- 一 統計學可分為敘述統計(Descriptive Statistics)及推論統計(Inferential Statistics)兩部份：

1. **敘述統計**：包含如何蒐集數據、展示數據、及找出可描述數據特徵之值的方法。
2. **推論統計**：包含如何由樣本資訊來推論群體，並估計該推論之可信度大小的方法。

- 解決統計問題之五大步驟：

步驟1：A clear **definition** of question to be answered and the **population** that relates to it.

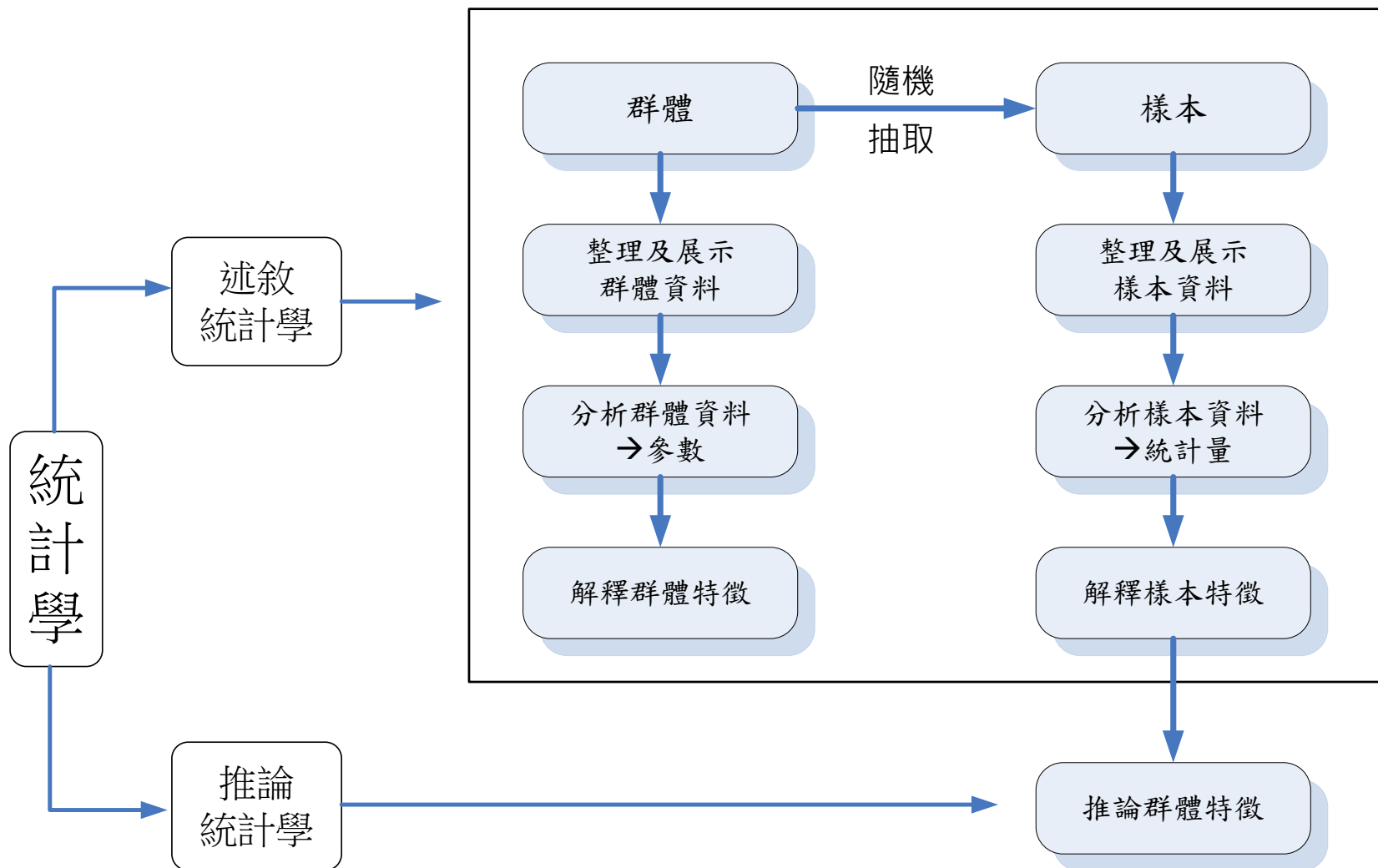
步驟2：The **design** of the experiment.

步驟3：The **collection** and **analysis** of data.

步驟4：The procedure for **making inferences**.

步驟5：The provision of a measure of “Goodness”(reliability) for the inference.

# Basic Statistical Concepts





- **Random Variable, (R.V ) (隨機變數)**

- A Survey statistician will most likely want to develop an instrument that asks several questions and deals with a variety of **phenomena** or **characteristics**. These phenomena or characteristics are called *random variable*. The **data**, which are the observed outcomes of these **random variables**, may differ from response to response. (研究者對所欲研究之群體所感興趣的一項或多項**特質**，稱為**隨機變數**)。

- **Qualitative data (定性資料)**

- Qualitative R.V. yields categorical data(類別資料). (即隨機變數的各結果不能以數量表示，而僅能依其特性之類別表之。)

如：1) gender性別

2) Nationality

3) Causes for product variation造成產品變異之可能原因

- **Quantitative R data (定量資料)**

- Quantitative R.V. yields numerical data(數值變數).  
(即隨機變數的各結果可以數量表之。)

1. **Discrete data(離散型資料)**

- 經由計數的方式取得資料。

2. **Continuous data(連續型資料)**

- 經由量測的方式取得資料。

# Types of data

- **Discrete data**

- 例：1) Defective items 不良品個數  
2) number of students in a classroom  
3) Defect counts on a wafer 晶圓上之缺陷點數

- **Continuous data**

- 例：1) weight  
2) height  
3) temperature

- 例 2 :

Determine the data type for the following R.Vs (決定下列隨機變數為定性或定量，若為定量則決定其屬離散型或連續型)。

- a) The defects found in a glass.
- b) Thickness of a wafer.
- c) Net weight of a pack of instant noodles.
- d) Possible causes for failing a Statistics class.

- Possible values for discrete data : 0 or positive integers ◦
- Possible values for continuous data : any real numbers.

# 常用統計圖表

條圖 (Bar Graph)、單圓圖 (Pie Chart) 與柏拉圖 (Pareto Diagram) 通常用來表示類別變數 (定性資料) 之次數分配。

- 條圖

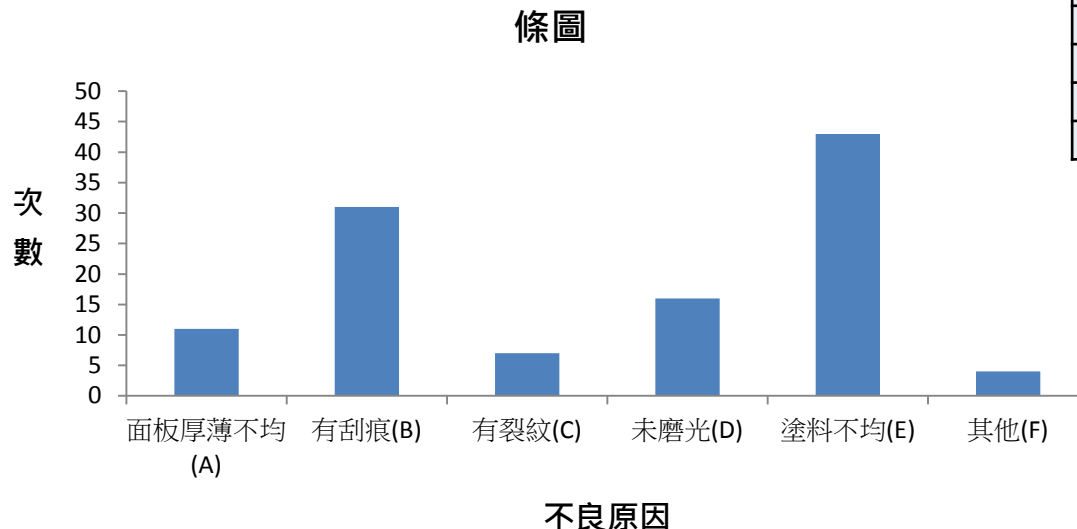
- 是用來比較及對照不同時期或類別間的差異。

# 常用統計圖表

## • 例 1：

- 一 某製造LCD面板的公司欲改善面板的品質，於是由該公司的品管人員收集了某天產量中不良品的資料，並列出造成不良品之原因的次數資料，如下表所示，請依數據繪製條圖。

不良原因	改善前
面板厚薄不均(A)	11
有刮痕(B)	31
有裂紋(C)	7
未磨光(D)	16
塗料不均(E)	43
其他(F)	4
總和	112



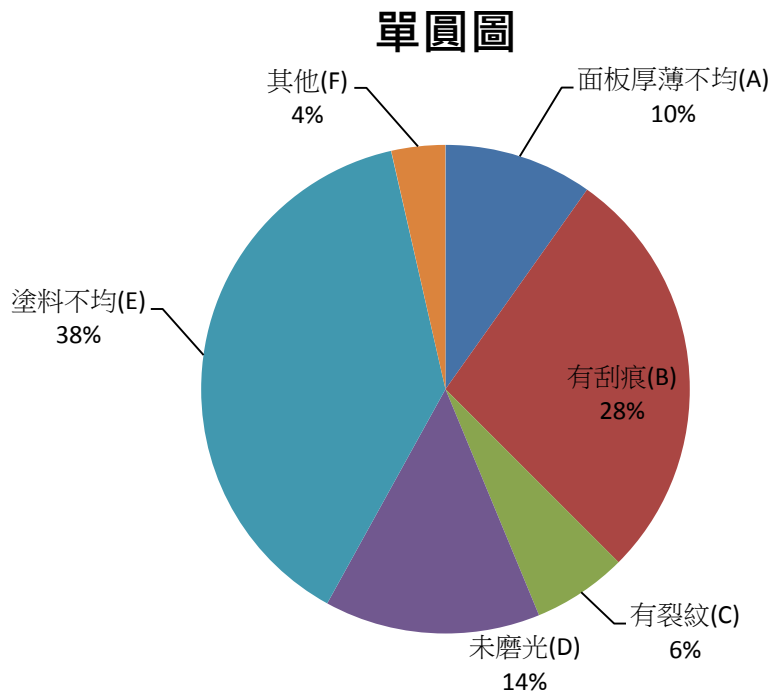


- 單圓圖

— 是用來顯示一個單一總合量如何攤分於各種類別中。

# 常用統計圖表

- 例 2：
  - 承例1，畫出單圓圖。



## • 柏拉圖

— 是定性（類別型）資料最常使用之圖形。

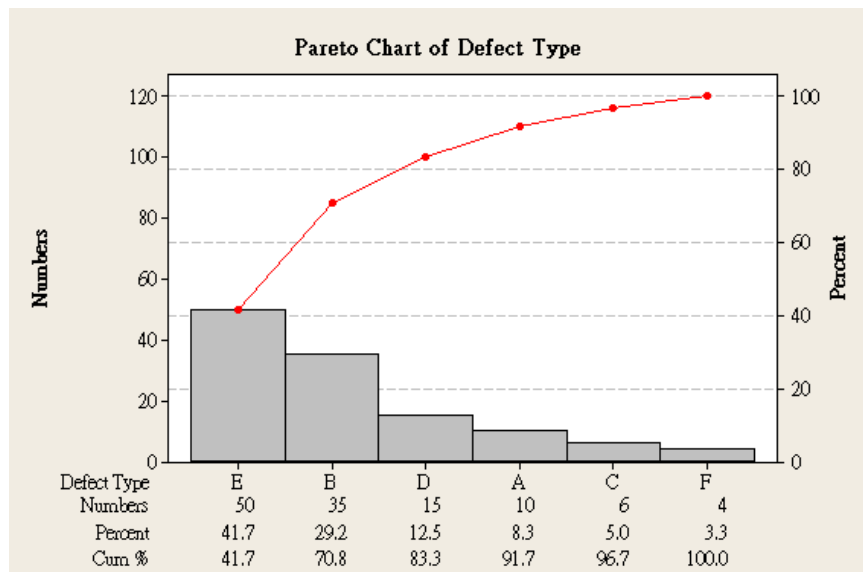
義大利經濟學家柏拉圖(Vilfredo Pareto)提出理論說明社會上大部份的財富均操縱在少數人的手中。在品質改善中，品質工程師也發現，造成問題或缺失的因素也符合柏拉圖原理。因此柏拉圖分析是依據「**重要少數，瑣細多數**」分類法則，進而找出造成問題（如：不良率過高）最關鍵之幾個少數因素。

# 常用統計圖表

## • 例3：

某工程師欲改善產品品質，收集了該產品在測試階段所發生之缺失資料，並將缺失的原因加以分類，列出缺失類型(Defect Type)的次數資料，如下表所示。試繪製柏拉圖。

Defect Type	Numbers
A	10
B	35
C	6
D	15
E	50
F	4
總和	120



- 時間序列圖

- 時間序列圖是用來表示資料在不同時間的關係圖，通常時間為橫軸，而縱軸則表示觀測值的單位數量。

## 常用統計圖表

- 例 4：

2013年1月至6月之美元兌歐元匯率：



資料來源：[www.exchangerates.org/history/USD/EUR/T](http://www.exchangerates.org/history/USD/EUR/T)

**莖葉圖 (Stem-and-Leaf Display) ，直方圖 (Histogram) 與散佈圖 (Scatter Diagram) 通常用來表示數值變數(定量資料)。**

- **莖葉圖**

- 每筆資料由莖(前頭的數字)與葉(最後一個數字)所快速組成。

- 例 5：

統計學教授隨機得到一組學生的期末考成績，請以莖葉圖畫出成績之分佈。

88 51 63 85 79 65 79 70 73 77



- 直方圖

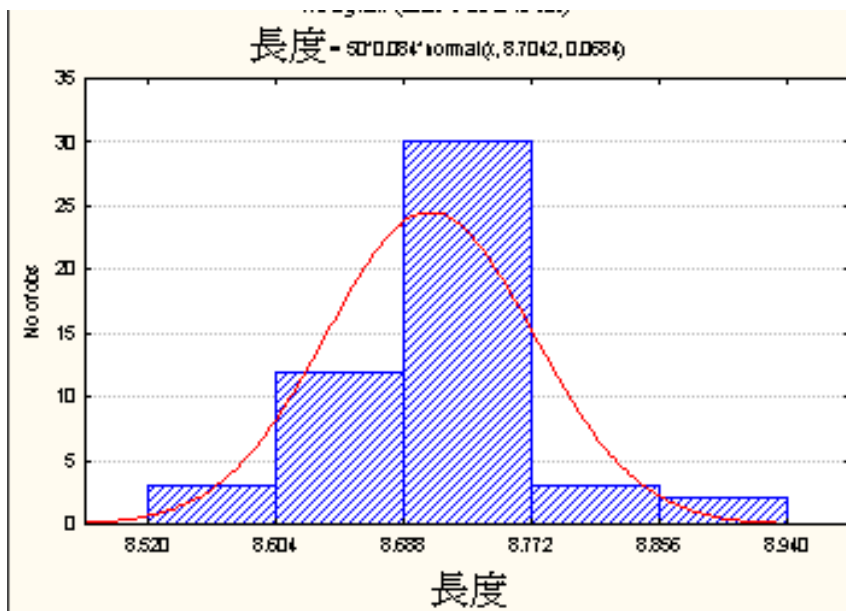
- 是連續型資料最常使用的圖形，用來展示資料之分佈。

# 常用統計圖表

## • 例 6：

某品質工程師隨機抽取其工廠生產之零件50個，分別測量其長度，得到數據如下：

試繪製零件長度之直方圖並分析之。



8.7	8.69	8.71	8.7	8.66
8.65	8.66	8.69	8.71	8.69
8.72	8.8	8.72	8.94	8.81
8.8	8.7	8.52	8.65	8.73
8.64	8.7	8.67	8.56	8.7
8.67	8.68	8.73	8.66	8.72
8.71	8.68	8.87	8.7	8.69
8.67	8.74	8.71	8.75	8.73
8.7	8.55	8.76	8.73	8.71
8.7	8.69	8.75	8.74	8.65

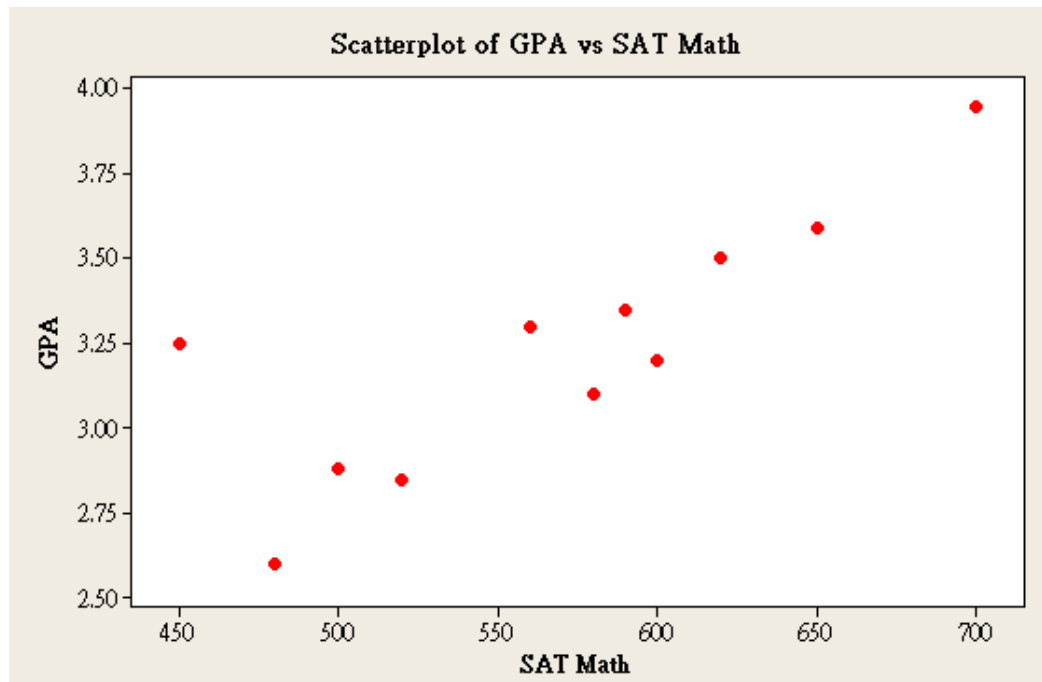
- 散佈圖

- 一 散佈圖是用來表示資料在兩個變數間的關係。

## 常用統計圖表

### • 例 7：

SAT數學成績對預測大學成績(GPA)而言，是否是一個重要指標？請利用下列SAT數學成績與GPA繪製散佈圖並分析之。假設得到數據如下：



SAT Math	GPA
450	3.25
480	2.6
500	2.88
520	2.85
560	3.3
580	3.1
590	3.35
600	3.2
620	3.5
650	3.59
700	3.95

# 數據之取得方式

- 普查

- 對群體中的每一個體取得資料，也就是100%的全檢。

- 抽樣 (Sampling)

- 利用一種程序或方法，由群體中抽出樣本。

- 常用的抽樣方法有下列四種：

- 簡單隨機抽樣
    - 系統抽樣
    - 分層隨機抽樣
    - 部落抽樣

## 簡單隨機抽樣

- 是指群體中每一個體被抽中之機會均等。
- 對群體內的每個個體編號，再以亂數表、電腦模擬亂數或製作紙籤的方法決定欲抽出之樣本。
- 優點：取樣方法簡便。
- 缺點：有時會因抽到的樣本過於集中某部分之群體，造成樣本之代表性不足。

## 系統抽樣

- 只做第一次隨機抽樣後，就依固定間隔數抽出一樣本。
- 例如：群體有30個樣本，預計抽出6個樣本，所以每隔5個即抽出一樣本。
- 優點：抽出第一個種子號碼，以後僅需每間隔數個樣本抽樣即可。
- 缺點：樣本在編號排序時必須與研究者所關心的變項無關，否則會造成樣本之代表性不足。

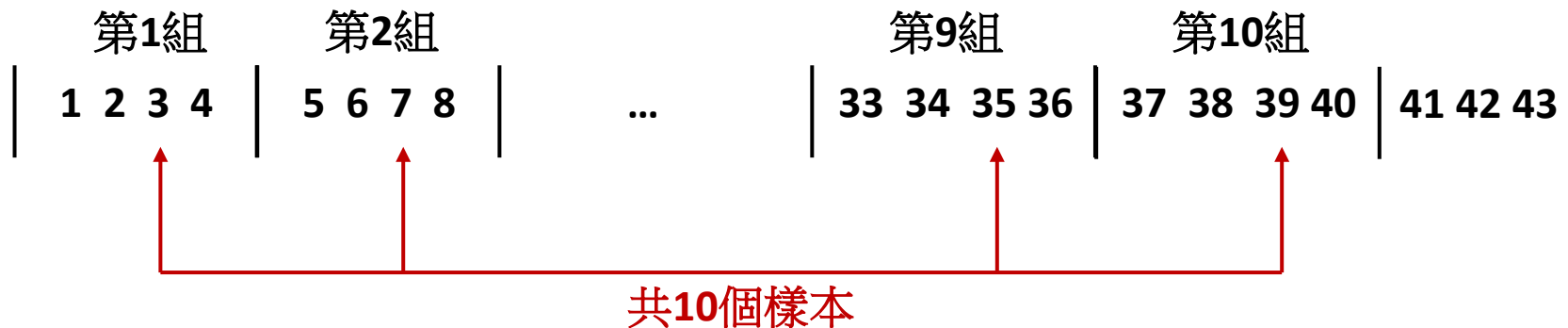


# 系統抽樣

## • 例：

假設群體含 43 個項目，預計抽取10個樣本，請以系統抽樣方法進行抽樣。

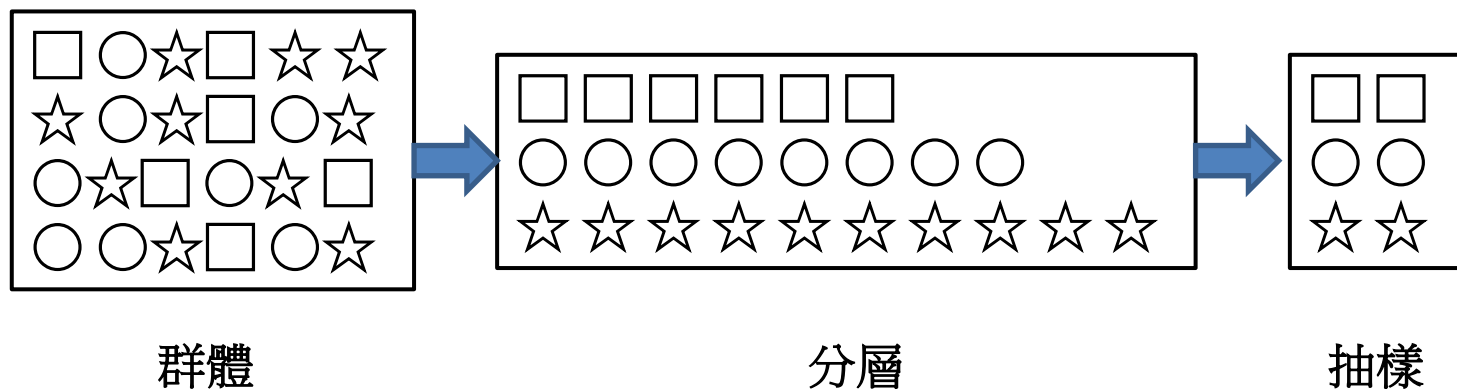
- 1) 編號
- 2)  $k = 4$  (取  $43/10$  之最大整數)，將所有43個項目，依順序每 4 個分一組。
- 3) 隨機抽取 1~4 中的一個數字，例如 3
- 4) 從編號3開始，間隔4個項目抽取一個樣本，即可獲得10個系統抽樣的樣本組



## 分層隨機抽樣

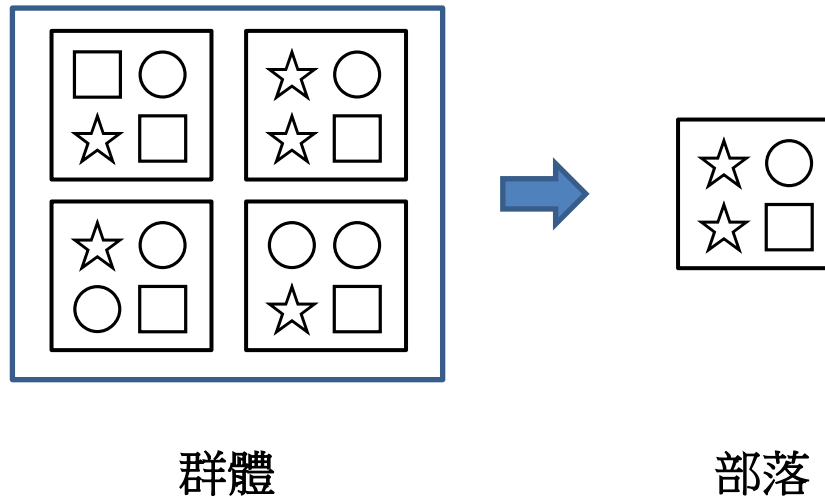
- 群體依某一衡量標準分成數個不重疊的子群（稱為層）。
- 將群體分層之後，再從每一層中利用簡單隨機方式取樣，即為分層隨機抽樣。
- 分層隨機抽樣之原則是同層內的性質差異要小，而不同層間之差異則要越大越好。

### • 例



## 部落抽樣

- 常用在群體中之個體分離很遠，且很難蒐集到樣本時。
- 部落抽樣是將群體先分成幾個部落，再從這些同一個部落中抽出一個或數個部落進行普查。
- 部落抽樣是假設每一個部落都是群體的縮影，因此不同部落間個體的性質差異性要小，而部落內個體的性質差異性要大。



# 數據之取得方式

## • 例 1：

交通大學某研究所共有100位研究生，某教授欲從此100位研究生抽10位，以估計研究生每日研讀論文的時數，將100位研究生依其學校排列如下：請問如何進行

- a) 簡單隨機抽樣
- b) 系統抽樣
- c) 分層隨機抽樣
- d) 部落抽樣

1	2	3	...	...	...	9	10
11	12	13	...	...	...	19	20
21	22	23	...	...	...	29	30
31	32	33	...	...	...	39	40
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
91	92	93	...	...	...	99	100

# 本單元結束