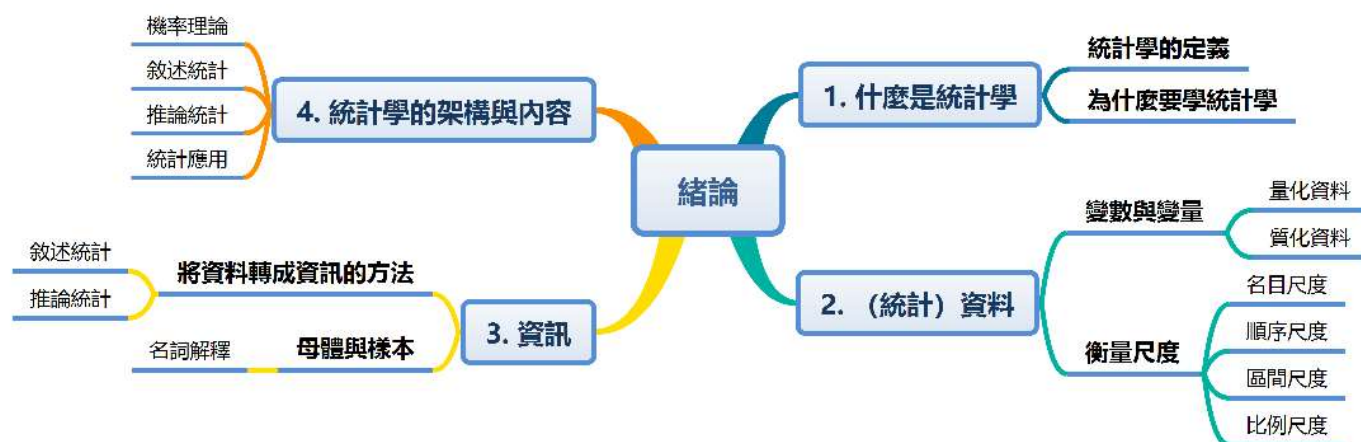


# 1 概論

>



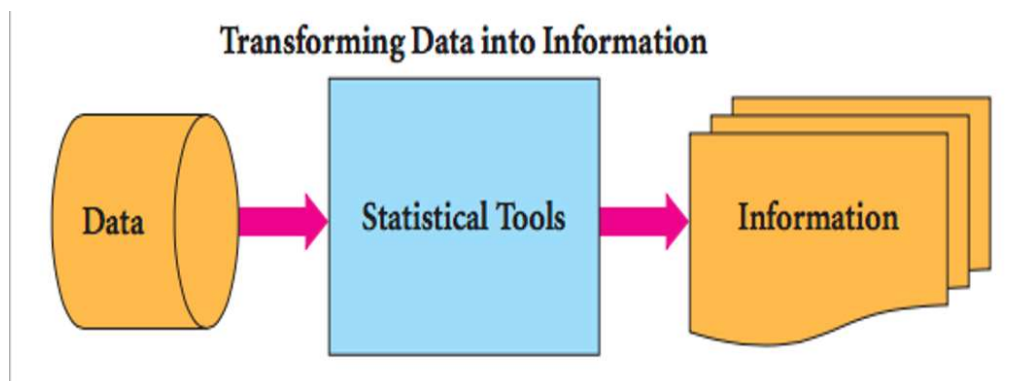
學習目標：

- (1)瞭解什麼是統計學？為什麼要學統計學？
- (2)瞭解資料的區分方法，包括質化資料與量化資料，以及名目、順序、區間與比例四種衡量尺度。
- (3)瞭解將資料轉成資訊的方法
- (4)瞭解統計學的基本架構

## 1-1. 什麼是統計學

### 1-1-1 統計學的定義

Def: 統計學是將資料轉成資訊的一門學問。



### 1-1-2 為什麼要學統計學？

運用統計學的觀念與方法，針對問題將研究對象的資料轉成資訊，幫助我們「以證據為基礎做決策」來解決問題。

例：電影魔球

<https://www.youtube.com/watch?v=EntXQjSdUMw>

### 1-1-3 統計學處理問題的流程



圖 1.2 統計學處理問題的流程圖

個案:灣景大學商管學生的道德行為

- Step1: 確定問題

2008年至2009年間的全球衰退中，有許多關於華爾街的主管、財務經理及其他公司主管的不道德指控。當時有些文章指出，此種不道德行為的部分原理可能源自商學院學生的作弊行為變得更普遍(Chronicle of Higher Education, Feb10,2009)。文章指出，56%的商學院學生承認在學校生活中作弊，比例高過非商學院的學生47%。

- Step2: 蒐集資料 (問卷調查)

這些年，灣景大學商學院院長一直在關注學生作弊行為。有些商學院同仁相信灣景大學的作弊行為比其他大學更普遍，但其他同仁則不認為作弊問題嚴重。因此，院長主導了一項研究，想要了解目前灣景大學商學院學生的道

德行為。研究的一部分是匿名的問卷調查，對象是畢業班的90名學生。學生被要求回答下列問題以取得三類作弊行為的相關資料。

- 1. 在灣景就讀期間，是否將抄襲自網際網路的資料當作自己的作品？ 是 \_ \_ \_ \_ 否 \_ \_ \_ \_
- 2. 在灣景就讀期間，是否在考試時抄襲其他同學的答案？ 是 \_ \_ \_ \_ 否 \_ \_ \_ \_
- 3. 在灣景就讀期間，是否與其他同學合作已完成應該自己獨立完成的作業？ 是 \_ \_ \_ \_ 否 \_ \_ \_ \_

回答一個(含)或以上的『是』的同學即被認為涉及某種形式的作弊行為。

部分資料如下，完整資料請參考名為Bayview的檔案。

Student	Copied from Internet	Copied on Exam	Collaborated on Individual Project	Gender
1	No	No	No	Female
2	No	No	No	Male
3	Yes	No	Yes	Male
4	Yes	Yes	No	Male
5	No	No	Yes	Male
6	Yes	No	No	Female
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
88	No	No	No	Male
89	No	Yes	Yes	Male
90	No	No	No	Female

### 管理報告

請為灣景大學的商學院院長準備一份報告，報告中彙整你對灣景大學商學院學生作弊行為的評估。

- Step3: 審查整理與呈現資料（敘述統計）
  - 1. 利用敘述統計量彙整這份資料。
- Step4: 分析解釋資料（推論統計）
  - 2. 求算男學生與女學生涉嫌某種形式的作弊行為比例為95%信賴區間估計值。
  - 3. 進行假設檢定，以判定灣景大學商學院學生作弊比例是否小於Chronicle of Higher Education所報導的商學院學生作弊比例。
- Step5: 統計推論（做決策）
  - 4. 根據你對資料分析，你會給商學院院長什麼建議？

## 1-2. 資料

學習統計學的第一步，就是要認識「資料data」，知道資料長什麼樣子，瞭解資料的基本型態，以及如何蒐集資料。

### 1-2-1 資料集

- 個體 ( individual) ，資料描述的主要對象。
- 變數(variable) ，又稱變項，用來描述特體的某種特性，這種特性是可以計量的，而變數的計量值或計數值稱為『變量』。
  - 例如：身高為一變數，其變量可能為50公分至200公分。
- 『資料』乃是由一個或多個變數之數值所組成的。

## Variables

Individuals

	Gender (M/F)	Age	Weight (lbs.)	Height (in.)	Smoking (1=No, 2=Yes)	Race
Patient #1	M	59	175	69	1	White
Patient #2	F	67	140	62	2	Black
Patient #3	F	73	155	59	1	Asian
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Patient #75	M	48	90	72	1	White

### 1-2-2 資料的分類

數據分析的邏輯不是依據數據的大小，而是依據數據的性質和數據間的邏輯關係。

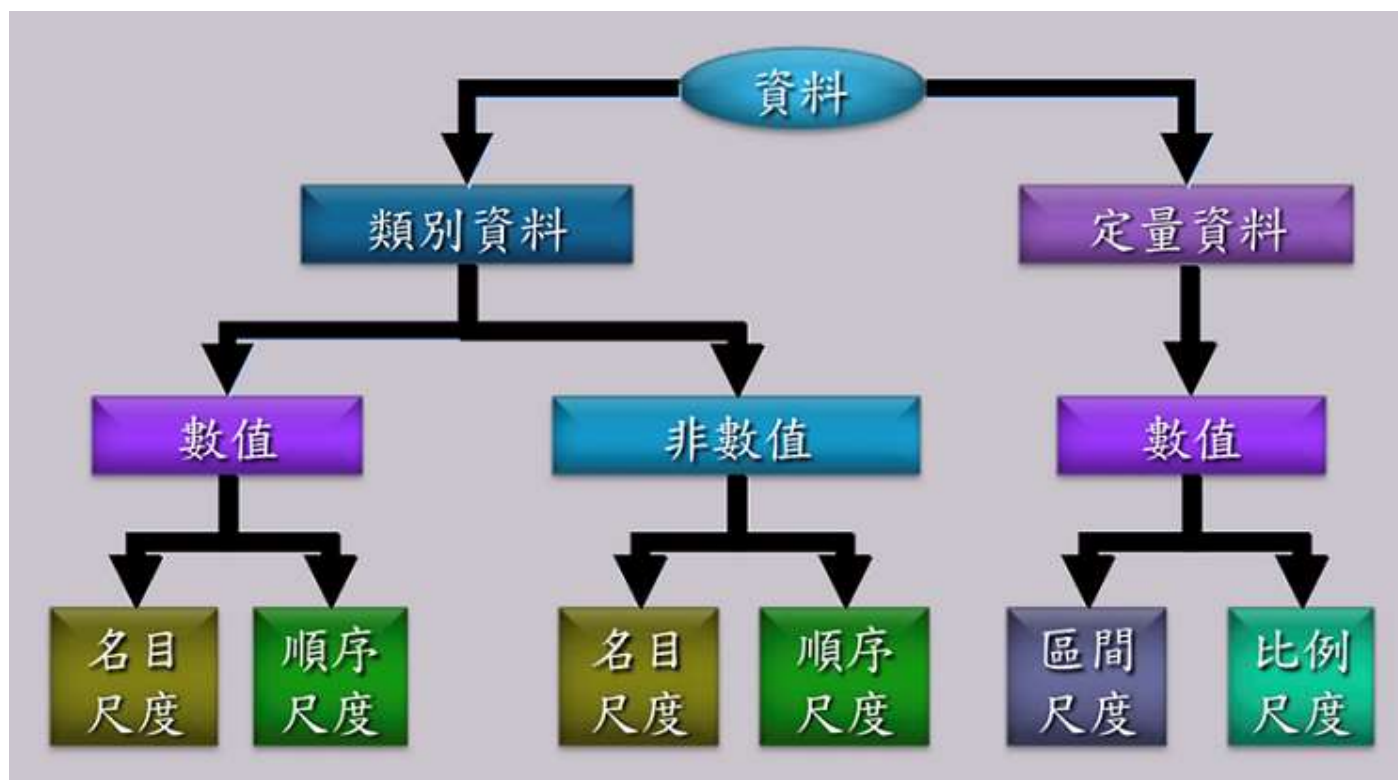
1. 資料依據變量的特性區分為『量化資料』(quantitative data)與『質化資料』(qualitative data)。
- 質化資料，例如性別、血型、婚姻狀況等分類性的資料。
  - 量化資料，是資料在數線上出現的狀況是連續的，例如身高、體重、商品售價、壓力、溫度等。

例: 試判下列資料的型態

1. A班學生中女學生的人數。
2. 某次國際電話的通話時間。
3. A公司之員工的教育程度。
4. 大一新生的宗教信仰。
5. 袋中彩球的顏色。
6. A員工一年內生病請假的天數。
7. 台灣地區一月份的降雨量。

解：計質資料：(3)、(4)、(5)。計量資料：(1)、(2)、(6)、(7)，其中離散資料為(1)、(6)，連續資料為(2)、(7)。

2. 資料依據衡量尺度分為『名目尺度』(nominal scale)、『順序尺度』(ordinal scale)、『區間尺度』(interval scale)與『比例尺度』(ratio scale)。



名目尺度的資料：

- 單純分類性的資料，可以簡單的數字或代號代表其屬性。
- 資料不具有順序性，只是代表類型。
- 例如性別、血型、婚姻狀況都是名目尺度資料。

順序尺度的資料：

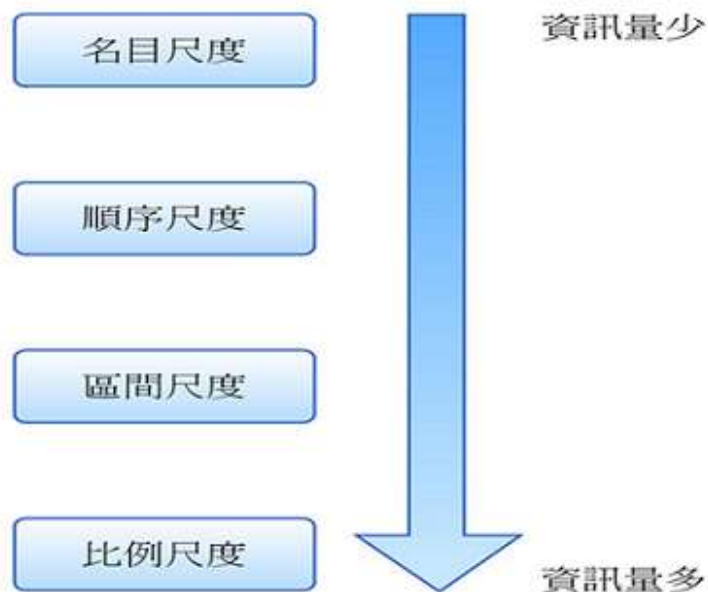
- 雖然是分類性的資料，但是具有順序性。
- 能夠排序，可依照資料的大小、強弱、好壞、重要性等程度區隔，但數值之間的差距不具意義。
- 例如教育程度、滿意度。
  - 教育程度可以依據受教育時間的長短，區分為小學以下、國中、高中、大專、研究所以上。

區間尺度的資料：

- 賦予研究對象數值，數值具有相對意義，可區分大小順序與距離，但沒有絕對的零。
- 數字與數字之間相加或相減有意義，相乘或相除沒有意義。
- 例如溫度、智商、時間。

比例尺度的資料：

- 賦予研究對象數值，數值具有絕對意義，可區分大小順序距離與比值。
- 資料不但能指出類別、排出順序、算出差距外，還帶有能夠求算數值之比例的資訊。
- 例如身高、薪水、小孩數目。

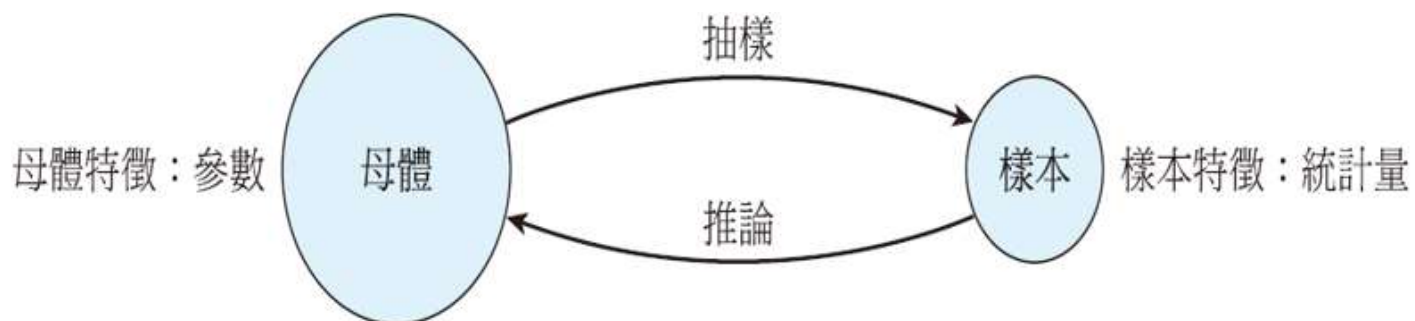


**圖 1.8** 各種資料尺度所含資訊量

## 1-3. 資訊

### 1-3-1 一葉知秋

瞭解現象（樣本），推論真相（母體）



#### 1. 母體(population)

所有研究對象的集合。

- 如果了解本校大學學生之體重情形，則世新大學全體學生所構成的集合，即為「母體」。

#### 2. 參數 (parameter)

描述母體特徵的數字，一般情況是未知的。

- 母體的平均數用來描述母體的平均狀況
  - 本校學生的平均體重，即為「參數」。

隨堂練習



某一研究者想瞭解全國失業率的問題，此時母體即為全體國民；同理，若他僅對台北市的失業率感興趣，則全體台北市民即成為其所欲研究的母體。由此可知，母體的範圍可大可小，完全視「所欲研究的對象」而定。

=> 台北市全體市民、台灣全體民眾

### 3. 樣本(sample)

部分研究對象的集合，由母體選取出的部份資料（子集合）。

- 從世新大學全體學生中所抽取出200名學生調查其體重，則這200名學生的體重所構成的集合，即為「樣本」。

### 4. 統計量(statistic)

使用樣本資料做計算，得出樣本的特性。

- 從大學學生中所抽取出200名學生的平均體重，即為一個「統計量」。

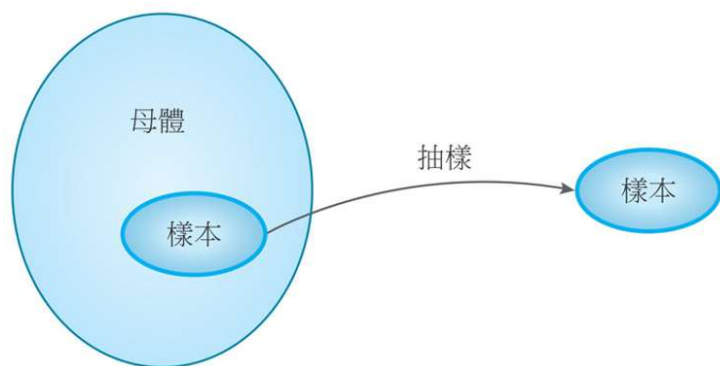


圖 1.3 抽樣示意圖

### 5. 抽樣(sampling)

若做為動詞解釋，是指由母體取出樣本的動作；做為名詞解釋，是指由母體取出樣本的程序或是方法。

- 從11,000名學生中取得2,000名學生的程序或過程，即為「抽樣」。

補充資料：

調查是蒐集資料常見的方法之一，所取得的資料是觀察性資料的一種。我們可以根據調查的範圍，將調查區分為普查和抽樣調查。

- 普查，是將整個母體納入觀察，以全部母體作為調查對象，也既是百分之百抽樣。例如每十年一次的戶口普查，每五年一次的工商及服務業普查、農林漁牧業普查。
- 抽樣調查，是以抽樣的方式從母體中抽取一部份個體作為樣本，然後就其調查結果推論母體的全部狀況。例如人力資源調查（失業調查）、物價調查、家庭收支調查、電視收視率調查、社會變遷基本調查、台灣選舉與民主化研究的抽樣調查、施政滿意度調查。

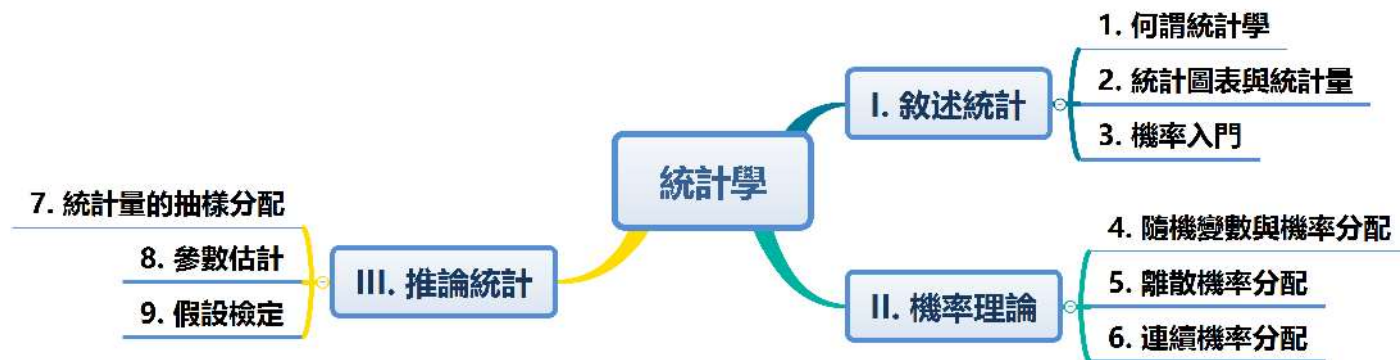
### Question: 為何統計學是數據分析之鑰？

任何數據分析都是以現象（或稱樣本）推論真相（或稱母體），事實的真相只有一個。數據分析的第一步是能分辨現象和真相，然後瞭解現象（敘述統計），並以已知的現象探索未知的真相（推論統計）。統計學的推論符合科學的方法，但是推論是有風險的，風險要以機率來描述。

## 1-4. 統計學的內容

### 1. 統計學的基本架構

對於資料的蒐集、整理、呈現、分析及解釋，並透過適當的機率模型，由抽取之樣本資料去推估母體未知之特性，希望能在不確定情況下，作成決策之科學方法稱之為統計學。



### 2. 將資料轉成資訊的方法

- 統計學分為敘述統計學與推論統計學兩大類型。
- 由敘述統計發展到推論統計，我們還需要有『機率理論』與『抽樣理論』為基礎。

#### (1) 『敘述統計』

透過統計表、統計圖以及各種量數，將原始資料的各種特性表達出來。

#### (2) 『推論統計』

利用抽出的樣本資料，經整理後去估計、檢定或預測母體中未知特性之科學方法。也就是從局部推論到全體的統計方法。

#### 估計 (estimation)

- 是指用樣本統計量去猜測母體參數的動作，例如用樣本平均數估計母體平均數。
- 點估計：以樣本資料計算所得之一個估計值，用來猜測母體參數。
- 區間估計：以樣本資料計算所得之一個範圍（區間），用來來猜測母體參數。

#### 假設檢定 (hypothesis testing)

- 是指先對母體參數做出假設，再用樣本資料去檢定假設是否成立。
- 銅板只有公正與不公正兩種，使用樣本資料來判斷銅板是公正還是不公正。

### 3. 架構間的關係



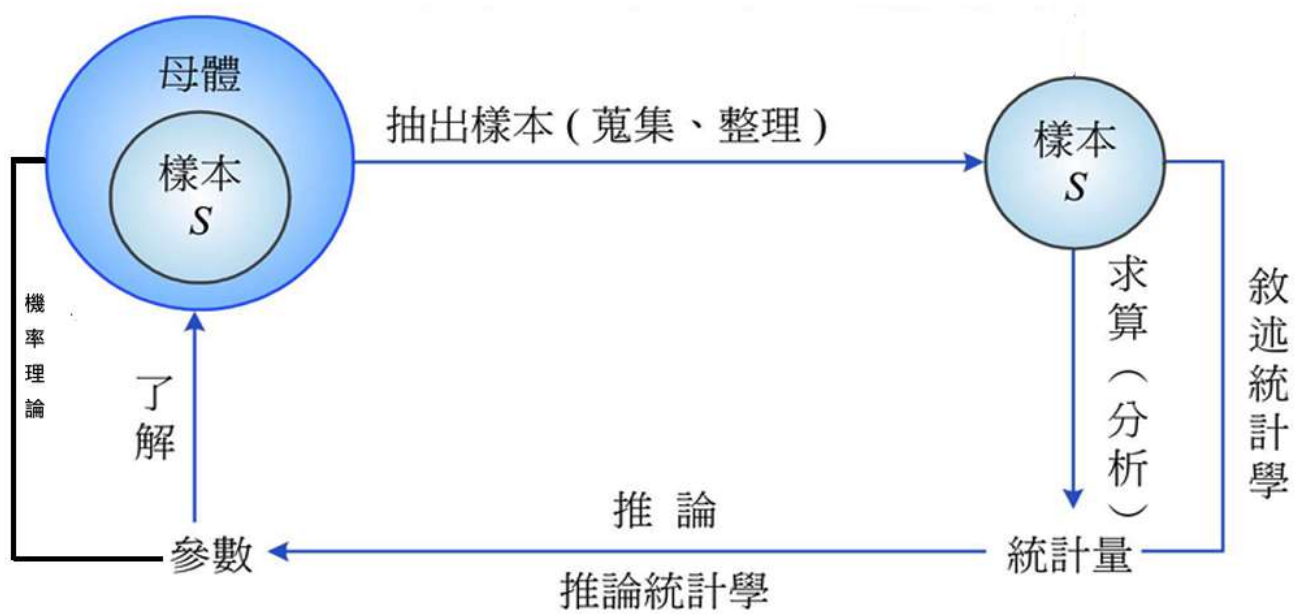


圖 1.7 統計學架構示意圖