

# 統計學講義

第一回

302211-1



社團法 考友社 出版發行

# 統計學講義 第一回



第一講 基本概念.....	1
命題大綱.....	1
重點整理.....	2
一、緒論.....	2
二、統計繪圖.....	11
三、敘述統計.....	28
精選試題.....	68

# 第一講 基本概念



## 一、緒論

- (一)統計學之意義
- (二)統計學之分類
- (三)母體與樣本
- (四)資料相關名詞
- (五)資料蒐集方法
- (六)衡量尺度
- (七)變數類別

## 二、統計繪圖

- (一)屬質性資料次數分配
- (二)屬量性資料次數分配
- (三)莖葉圖
- (四)長條圖
- (五)圓餅圖
- (六)直方圖
- (七)折線圖
- (八)肩形圖

## 三、敘述統計

- (一)概述
- (二)中央傾向
- (三)離散度
- (四)偏態與峰態
- (五)百分等位與百分位序
- (六)盒鬚圖

\*\*\*\*\*  
\*  
\* **重點整理** \*  
\*  
\*\*\*\*\*

## 一、緒論

### (一)統計學之意義：

1. 現今人類所面對的各種現象已愈趨複雜多樣、雜亂紛陳，爲了從這些大量、紛亂、具有不確定性的現象裡，找出其中有用的訊息與原則，一種專爲處理不確定問題的科學方法便應運而生，這就是所謂的統計學（Statistics）。
2. 簡而言之，統計學就是一種有系統的科學方法與研究工具。它運用蒐集、整理、分析、解釋資料等程序，將其轉化爲簡潔有序、脈絡分明的資訊；同時藉由嚴謹、科學化的推論與預測，可在不確定狀況下做出明智的決策。
3. 目前統計學已廣泛應用在許多領域，如個人或家庭的收支預算或保險理財、企業的經營決策或產品定位、國內外經濟問題的研究、政府公共政策的釐定、全球人口成長的結構分析…等。由此可見，統計學並非針對某一特定領域探究，而是一種可用於許多學門的研究方法。
4. 其他像是溫室效應與聖嬰現象對全球造成的影響爲何？人民對政府官員的滿意度爲何？調漲大學學費是否會影響升學意願？投資股票是否會賺錢？抽菸、嚼檳榔和癌症之間是否有直接關聯？核四廠的興建是否真的有其迫切性？…等，也都可以透過統計的方法來分析、解釋、甚至預測。
5. 雖說統計學的應用領域和應用方法不斷地擴展，但統計方法的應用並非毫無限制。茲歸納應注意的兩點如下：
  - (1)資料應正確。
  - (2)方法應正確。

### (二)統計學之分類：

#### 1. 依分析順序：

##### (1)敘述統計（Descriptive Statistics）：

又被稱爲資料分析（Data Analysis），是統計方法的初步程序。其主要的目的是將原始資料作一番整理、組織和表述，並且利用各種統計圖表及統計量的計算，使統計的結果清楚展現出來。這部分的過程相

當繁瑣，但卻是後續分析、推論最至關緊要的基礎。

(2)推論統計 (Inferential Statistics) :

主要目的則在於由樣本資料推估母體參數，以作為策略擬定之決策依據。例如，飲料廠商想瞭解所開發的產品甜度是否適中，則可藉由適當的抽樣，抽取代表性的消費者，進行口味測試，並進而推估全市場對甜度接受的情況。

2. 依變量多寡：

(1)單變量統計 (Univariate Statistics) :

探討的變量只有一個，是統計學領域中最根本、也最易理解的基礎分析，如單因子變異數分析 (One-Way ANOVA)、簡單迴歸分析、複迴歸分析、時間序列分析…等。

(2)多變量統計 (Multivariate Statistics) :

為單變量統計的延伸，所探討的變量則有兩個或兩個以上，如多變量迴歸分析、多變量變異數分析 (MANOVA)、鑑別分析、主成份分析、因素分析、集群分析、多元尺度分析…等。

3. 依研究範疇：

(1)理論統計 (Theoretical Statistics) :

為統計的學理基礎，著重在公式的推導及定理的證明，故較偏向數理統計方面的討論，須善加活用微積分、線性代數…等工具。例如，貝氏定理、常態分配理論、中央極限定理…等，均涉及較深入的數學推演，屬理論統計的範疇。

(2)應用統計 (Applied Statistics) :

用以解決真實世界的諸多問題，其目的在善用理論統計中的各種定理或公式，支援決策之制定。例如，變異數分析、迴歸分析、時間序列預測…等，均以理論統計為基礎，應用到諸多研究領域。以企業管理而言，不論作業管理、行銷管理、人資管理、資訊管理、財務管理，只要稍微深入的分析，都離開不了統計的應用。

4. 依母體假定：

(1)母數統計 (Parametric Statistics) :

在統計推論前，須先假定母體的分配型態。一般而言，母數統計適合用在像社會科學、管理科學等大樣本的研究。統計學領域中，幾乎大部分的探討主題都是母數統計。例如，單母體假說檢定、雙母體假說檢定、變異數分析…等，均假定統計量遵循常態分配，為典型的母數統計。

(2)無母數統計 (Non-Parametric Statistics) :

無須假定母體的分配型態。一般而言，無母數統計適合用於生命科

學的小樣本研究。許多的研究由於樣本數少，運用常態假定並不適合。例如，卡方檢定、符號檢定、等級相關檢定…等，在統計的推論過程中，均未假定資料遵循常態分配，為典型的無母數統計。

#### 5. 依學習順序：

##### (1) 機率論 (Probability Theory)：

為統計推論的基礎，主要在探討各項機率概念，並藉之深入隨機變數與抽樣分配之理論。機率論探討聯合機率、邊際機率、條件機率等問題，並深入間斷性隨機變數、連續性隨機變數之分配。

##### (2) 統計推論 (Statistical Inference)：

以所觀察的樣本，運用機率論，對母體參數進行估計或檢定，故統計推論可說是機率論的延伸。此處統計推論與前述推論統計兩者內容的差異其實有限，只是前者較適合與機率論並列，而後者較適合與敘述統計並列。

#### 6. 依時空斷面：

##### (1) 橫斷面分析 (Cross-Sectional Analysis)：

探討特定時點下的結構，其性質猶如雷達般地偵測 (Scanning) 現況。常見的橫斷面分析有消費者調查、民意調查、犯罪心理、生態調查等研究。

##### (2) 縱斷面分析 (Longitudinal Analysis)：

探討跨時點的結構，其性質猶如水晶球般地預測 (Forecasting) 未來。縱斷面分析大多應用在經濟、產業、市場的研究上，包括經濟預測、股票分析、市場需求預測等。

#### (三) 母體與樣本：

母體的範圍相當龐大，且難明確界定，應用上須藉由抽樣 (Sampling) 自母體中抽取出部分資料進行分析，所抽取出來的資料即為樣本。由樣本求算出統計量 (Statistics)，統計量再經由統計推論，即可估得母體參數 (Population Parameters)，母體參數則可用以描述母體特性。

##### 1. 母體：

(1) 所謂母體 (Population)，指的是研究對象所有個體的集合，亦即機率論中所稱的宇集 (Universal Set)。母體中的個體可以是個人、家庭、公司、國家…等，端視研究目的而定。母體的數量多半相當龐大，有些甚至是觀察不到或無法確定的。例如，乳品廠商欲估計每人牛乳的需求量，個體就是「人」；又如欲對前 500 大製造業的年度生產年營業額進行調查，則其個體就是「公司」。

(2) 母體範圍的界定是非常重要的。例如，欲探討設籍臺北市之國中生升學進入公立高中的情況，則研究對象 (母體) 就是「設籍於臺北

市的國中生」，而非臺北市內所有的國中學生。因為就讀國中的學生中，有部分學生是從其他縣市跨區就讀的，因此必須將非設籍於臺北市的國中學生加以排除。

- (3)母體範圍界定錯誤，會導致樣本抽取錯誤，進而導致後續的統計分析、檢定、推論等都將產生與事實不符的結果。因此母體的界定必須清楚正確，符合所欲研究的對象，才不會產生誤導的結果。

## 2. 樣本：

- (1)所謂樣本（Sample），指的是母體中部分個體的集合，亦即機率論中所謂的子集（Subset），故樣本來自於母體。由於母體經常是非常龐大且難以全面觀察的，因而有必要自母體中選取部分個體加以觀察，再由這些樣本資料經統計方法推求母體的特質。
- (2)例如，民意調查機構想調查大學生對於當今政壇的看法，由於全國大學生人數甚多，逐一訪談勢不可行。因此通常會選取一部分的大學生作為樣本進行問卷調查，以推測全體大學生對這些問題的看法。
- (3)樣本對母體的代表性有賴於正確的抽樣方法與足夠的樣本數目。否則縱使母體的界定極為正確，只有抽樣方法運用不當，所抽取的樣本就不可用以推論母體；同時，若樣本大小不及一定的標準，亦不允許進行統計推論，以免產生嚴重的誤導。

## 3. 普查與抽查：

- (1)研究者針對母體做全面性的觀察，即母體中所有個體皆在調查範圍內，稱為普查（Census）。例如，政府每 10 年都會由各戶籍所在地的警察機關挨家挨戶地進行人口數調查，以正確統計全國實際人口數，這就是一種普查（人口普查）。
- (2)反之，若只從母體中抽取部分樣本作為調查標的，則稱為抽查（Survey）。例如，工廠每天生產數萬個燈泡，只選取其中幾盒來檢驗不良品的比率，這就是一種抽查。
- (3)在實務上，普查極為罕見，因為母體通常極為龐大、甚至無法加以觀察。普查對母體具有高度破壞性，如燈泡壽命的檢查往往必須要等到燈絲燒斷，才能知道結果。而抽查只要方法適當，就能選取到足夠多的有效樣本對母體進行推論，實為省時、省力的研究方法。
- (4)況且若所有的問題都可使用普查來解決，就不需要由樣本來推估母體了。換言之，此時統計學便沒有任何存在的價值。今日統計學之得以蓬勃發展，抽樣理論的貢獻甚大。

## 4. 隨機樣本與非隨機樣本：

- (1)由母體所抽取出的樣本，依抽樣方法之不同可分為隨機樣本與非隨機樣本。抽樣方法又可分為隨機抽樣與非隨機抽樣。

- (2)由隨機抽樣所獲得之樣本，稱為隨機樣本（Random Sample），較能代表母體。由非隨機抽樣所獲得之樣本，稱為非隨機樣本（Non-random Sample），較不能代表母體。一般而言，隨機樣本的取得成本較高，在統計推論上也較有價值；但經常礙於成本的考量，會以非隨機樣本取代隨機樣本。
- (3)隨機抽樣在抽取樣本之前，每一個體被抽取的機率相等；而非隨機抽樣在抽取樣本之前，每一個體被抽取的機率則不相等。例如，女性保養霜公司想瞭解市場對新產品的接受度，母體為所有的成年女性；但由於成本的考量，未赴市場抽樣，僅以公司的女性員工為樣本，進行調查，則所抽取的樣本為非隨機樣本。例如，若一個紙盒中有 50 顆球，每次只能取出一顆球，則每次取出後再放回紙盒的抽樣方式就是隨機抽樣，所抽出來的球則是隨機樣本；反之，若取出後不再置回，則為非隨機抽樣，其樣本稱為非隨機樣本。
- (4)隨機抽樣純粹依機率原則自母體中抽取樣本，不因個體之特性、人為意志或其他因素而有所不同，如此所抽取的樣本便稱為隨機樣本。反之，若母體中各個體被抽取的機會不等，某些個體被抽到的可能性遠高於其他個體，則這個過程就是為非隨機抽樣，而其所抽取的樣本就稱為非隨機樣本。
- (5)由於非隨機樣本不能公正客觀地代表母體特性，統計學上能用以推論的樣本亦必須為隨機樣本，因此一般初等統計所討論的抽樣理論皆限於隨機抽樣，以期使樣本推論母體能得到最精確的結果。

(四)資料相關名詞：

當資料蒐集完成後，為方便觀察和計算起見，通常會予以編製成資料檔，如下表所示。表中包含實體、變數、觀察值、個案等，如此可將所有的原始資料檔結合起來並清楚地呈現出來。資料不但是統計學的基礎，也是計算機科學的基本元素。

員工姓名	部門	職稱	年齡	性別	變數
趙國棟	行銷部	專員	40	男	← 實體
王美華	營業部	業務代表	28	女	→ 個案
謝淑霞	會計部	會計	35	女	→ 觀察值
張凱立	資訊部	工程師	32	男	

↓  
資料檔

南海公司員工資料表



## 1. 實體：

所謂實體（Entity），指的是資料中具有某些特徵的個體，可以是個人、家庭、公司、國家…等，端視資料的性質而定。例如「南海公司員工資料表」中「趙國棟」即為一實體。

## 2. 變數：

所謂變數（Variable），指的是代表實體特質的一些項目，一份資料檔中可有許多個不同的變數。例如「南海公司員工資料表」中「性別」、「年齡」皆是該資料檔中的變數。在計算機科學中，則稱為欄位（Field）或縱行（Column）。

## 3. 觀察值：

觀察值（Observation）是一實體在特定變數之數值。例如「南海公司員工資料表」中，員工姓名「張凱立」，對應的變數為「年齡」，則兩者交會所得之「32」即是一個觀察值。計算機科學中，該項目則稱為資料值（Data Value）。

## 4. 個案：

個案（Case）為一實體與所有變數對應的觀察值集合。例如「南海公司員工資料表」中，員工姓名「王美華」，對應的變數有「年齡」、「性別」，這一橫列就是一個個案。計算機科學中，則稱為橫列（Row）或紀錄（Record）。

## 5. 資料檔：

資料檔（Data File）即是所有觀察值的集合，亦即上述各項目的集合。例如「南海公司員工資料表」整體便是一份資料檔，內含許多的觀察值。計算機科學中，則稱為資料檔（Data File）或資料表（Data Table）。

## (五)資料蒐集方法：

統計的基礎為資料，而要獲取精確的資料則有賴於正確的蒐集方法。在探討資料蒐集方法（Methods of Data Collection）之前，必須對資料的類別先有概括性的認知。資料在來源上可分為初級資料，或稱為一手資料，以及次級資料，也就是所謂的二手資料。兩者的差異在於初級資料必須經由研究者親自蒐集才能獲得，如問卷調查、電話訪查、實驗設計…等；而次級資料則係利用現成之既有資料，如政府統計、研究報告、傳播媒體…等。不同的資料類別在蒐集的方法上有不盡相同的蒐集方式。

## 1. 初級資料：

初級資料（Primary Data）指非現可取得的資料，需要研究者新蒐集

♥♥♥♥♥♥♥♥♥♥♥♥♥♥♥♥  
♥  
♥ **精選試題** ♥  
♥♥♥♥♥♥♥♥♥♥♥♥♥♥♥♥

一、下表為甲班的基本資料和數學期中考成績：

姓 名	學號	性別	期中考成績
丁大維	10	男	95
王文中	25	男	68
張海婷	33	女	43
羅書豪	39	男	71
陳玉芬	51	女	82

- (一)試找出這個資料檔的變數。  
 (二)任舉一觀察值，並說明其意義。

答：(一)變數為「姓名」、「學號」、「性別」、「期中考成績」。

- (二)選取「丁大維」和「期中考成績」兩者的交集，得到「95」，95 就是一個觀察值，它代表了丁大維在這次數學期中考得到 95 分。

二、一筆資料中取出 5 個觀察值，分別為 1.3、7.0、3.6、4.1、5.0，試回答下列各題：

- (一)算數平均數為何？  
 (二)說明(一)所得之平均數為該筆資料的平衡點。

(三)以上列觀察值說明  $\sum_{i=1}^n (X_i - \bar{X}) = 0$ 。

(四)以上列觀察值印證  $\sum_{i=1}^n (X_i - \bar{X})^2 < \sum_{i=1}^n (X_i - A)^2$ ， $A \neq \bar{X}$ 。

答：(一)算數平均數  $\bar{X} = \frac{\sum_{i=1}^k X_i}{n} = \frac{1.3 + 7.0 + 3.6 + 4.1 + 5.0}{5} = 4.2$

- (二) 1. 比 4.2 小的數值有 1.3、3.6、4.1，距離和為：

$$|1.3 - 4.2| + |3.6 - 4.2| + |4.1 - 4.2| = 3.6$$

2. 比 4.2 大的數值有 7.0、5.0，距離和為：

$$|7.0-4.2|+|5.0-4.2|=3.6$$

3. 故知算數平均數是整個資料的平衡點。

(三)(1.3-4.2)+(7.0-4.2)+(3.6-4.2)+(4.1-4.2)+(5.0-4.2)=0，故得：

$$1. \sum_{i=1}^n (X_i - \bar{X}) = 0$$

$$\begin{aligned} 2. \sum_{i=1}^n (X_i - \bar{X})^2 \\ = (1.3-4.2)^2 + (7.0-4.2)^2 + (3.6-4.2)^2 + (4.1-4.2)^2 + (5.0-4.2)^2 \\ = 17.26 \end{aligned}$$

(四) 1. 令 A 為不等於 4.2 之任意數，如 A=4.1，則：

$$\sum_{i=1}^n (X_i - A)^2 = 17.31$$

2. 故可印證：

$$\sum_{i=1}^n (X_i - \bar{X})^2 < \sum_{i=1}^n (X_i - A)^2$$

三、甲、乙、丙三班之學生人數分別為 50、55、48 人。設三個班級的平均體重分別為 60、56、59 公斤，試求三個班級的總平均體重。

答：(一)這是好幾組資料在給定各組平均數的情況下，求共同平均數。先計算出每班的總體重後相加，再除以三個班級的總人數，即：

$$\begin{aligned} \mu &= \frac{n_1 \cdot \mu_1 + n_2 \cdot \mu_2 + n_3 \cdot \mu_3}{n_1 + n_2 + n_3} = \frac{50}{153} \times 60 + \frac{55}{153} \times 56 + \frac{48}{153} \times 59 \\ &= 58.248 \end{aligned}$$

(二)其中，

$$W_1 = \frac{50}{153}, W_2 = \frac{55}{153}, W_3 = \frac{48}{153}$$

(三)故這三個班級的平均體重是 58.248 公斤。

四、假定某院學生分為甲、乙、丙三組上微積分課，甲組有學生 40 人，平均成績 80 分，標準差 8 分；乙組有學生 30 人，平均成績 75 分，標準