

第一章

敘述統計概論

大綱

- ◎ 什麼是統計學
- ◎ 統計學的分類
- ◎ 資料蒐集與分類
- ◎ 資料的度量尺度
- ◎ 定量資料的統計圖表
- ◎ 集中趨勢量數
- ◎ 中央趨勢度量（單峰分配）
- ◎ 離散程度衡量
- ◎ 偏態係數與峰態係數
- ◎ 觀測值標準化與經驗法則





內文教學區

一、什麼是統計學

(一) 統計的意義

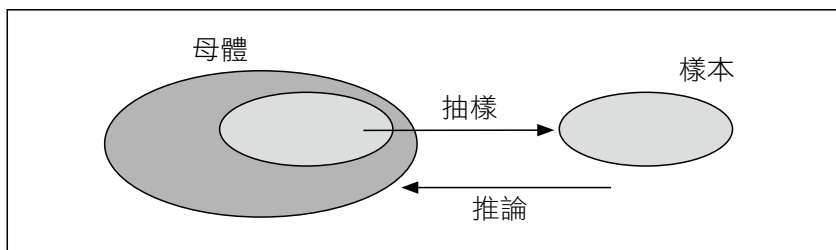
統計係利用數據變動的資訊及經過資料蒐集和整理、表現、分析和解釋，再依據分析出來的結果，作出較為正確且合理的決策，並大幅降低結果之不確定性的科學方法。

※簡言之，統計就是將大量的資料轉化成有用的資訊。

(二) 統計學常用名詞解釋

1. 母體：人們所要研究的對象全體稱為母體（population），它是一群具有某種共同特性的人或事物之個體所構成的群體，而每一群體通常含有多種特性，可依某種特性歸納為一母體，亦可依另一種特性歸納為另一母體。例如，臺大醫院想了解 99 年度該院新生兒之平均體重，則 99 年度臺大醫院所有新生兒所成的集合，為此研究的母體。
2. 樣本：樣本由抽自母體中之某些個體測量所構成。例如，從 99 年度臺大醫院新生兒中隨機抽出 200 位，則此 200 位新生兒所成的集合，為此研究的樣本。

母體與樣本的關係如下圖所示：



母體與樣本之關係圖

3. 母體參數：用來代表母體特徵的量值。例如，99 年度臺大醫院所有新生兒之平均體重（稱為母體平均數 μ ），為此處所謂的參數。
4. 樣本統計量：用來代表樣本特徵的量值。例如，從 99 年度臺大醫院新生兒中隨機抽出 200 位之平均體重（稱為樣本平均數 $\bar{X}$ ），為此處所謂的統計量。

- 【觀念題 1】

【解】(A)

(1) 觀察：科學家蒐集、研究與其問題有關的事實，因而必須觀察研究對象發生的情形。



- (2) 假說：科學家解釋所觀察的事實，然後將其預感或直覺轉化為假說或理論，以表達洞察資料的結果。
 - (3) 預測：科學家從假說及其相關理論中推演出某些結果。若假說成立，則此結果是從假說得出的新知識，但卻不是從經驗中所獲得的新知識。若所得理論有價值或合理，則它可能會成為一種新知識。這些新事實通常稱為預言，是還未作觀察前之預言。
 - (4) 驗證：科學家蒐集新事實（即觀察）去檢驗由假說或理論所推測的結果是否成立。這個步驟使上述三個步驟必須重複進行。若理論被證實，則由此作更明確、更深入的預測，然後再度檢驗這些預測，直至發現某些偏差而修訂理論。若與理論矛盾，則可明確陳述與眾多事實相符合的新假設，然後再重新進行第三個、第四個步驟。
2. 統計學與科學方法的關聯：
- (1) 統計學主要是與第一個及第四個步驟（即觀察與驗證）有密切的關係，且與第三個步驟（預測）有某種程度的關聯。
 - (2) 近代統計學的推論過程，已給予假設的檢驗更明確的意義。統計學上的「檢驗」假設，也許「檢定」是一個更好的字眼，直至機率在證據上的影響能夠正確給予解釋的程度。現代的統計方法已經能夠以數值機率客觀地度量不正確解釋的危險值。
3. 演繹法與歸納法：各種科學之間是用科學方法連貫起來的，而在這些方法中包括了演繹法跟歸納法。演繹法就是數學，歸納法就是統計學。要應用演繹法或歸納法便少不了要做數字上的紀錄，而那些將觀測現象變成數值的設施即是科學儀器。

【觀念題 2】

下列調查結果何者屬於敘述統計？

- (A) 由於香蕉生產過剩，預計今年香蕉的平均價格將跌到每台斤不到 5 元台幣。
- (B) 由於政府施行禁菸規定，今年二月份的香菸銷售量較去年同期減少了 5%。
- (C) 由於全球金融風暴，紐約地區今年失業率將上升 6%。
- (D) 由於近年少子化的影響，預計自 2010 年開始，臺灣的人口將呈現負成長。

【解】(B)

**【觀念題 3】**

下列調查結果何者屬於推論統計？

- (A) 因應政府拯救房市方案，今年初各大銀行紛紛調降房屋貸款利率，平均利率下降了 2 個百分點。
- (B) 由於政府去年鐵腕施行掃毒行動，今年上半年的犯罪率較去年同期減少了 5 個百分點。
- (C) 由於臺灣地區鰻魚生產過剩，預計今年外銷鰻魚的平均價格將跌到每台斤不到 100 元台幣。
- (D) 由於全球金融風暴，臺灣地區今年一月份的失業率較去年同期上升了 0.6 個百分點。

【解】(C)

三、資料蒐集與分類

(一) 統計工作程序

辦理統計工作的程序，包括：事前計劃、蒐集資料、整理資料、分析資料、資料表現等。

(二) 資料分類

統計資料的種類繁多，按資料的性質，可分為質性資料與量化資料；按資料的來源，可分為初級資料與次級資料。

1. 質化資料：其特性為不能衡量、以點計算的，以分類或排序為主的資料。如色澤、職業等。

2. 量化資料：其特性可計點、衡量的，以數值或測度為主的資料。又分為：

(1) 離散型資料 (discrete data)：可數的 (countable) 數值資料。例如，人數。

(2) 連續型資料 (continuous data)：不可數的 (uncountable) 的數值資料。例如，長度、重量、時間。

※ 欲判斷何時用何種統計分析方法，第一步要判定各個變數（不論是自變數或因變數）的性質是離散的 (discrete) 或連續性的 (continuous)。

3. 初級資料：係指研究者直接由調查或試驗而蒐集得來的資料，亦稱原始資料或直接資料。例如，民意調查機構要預測選舉結果，而直接向選民詢問調查所得的資料。初級資料蒐集不易，有賴



於完善的蒐集技術與籌劃，雖然其須花費大量的人力、時間、金錢，但是蒐集來的資料較為周詳而合用。

4. 次級資料：係指其他私人、機關、團體所蒐集整理後的資料，亦稱為現成資料或間接資料。

(1) 次級資料的來源：

- a. 政府機關所編製的統計報告。例如，臺灣人口統計年報、進出口貿易統計月報、臺北市物價統計月刊等。
- b. 企業機構所發行的統計刊物。例如，報紙上的證券行情表、物價行情表等。
- c. 學術團體所出版的書刊雜誌。例如，中國統計學報、各校的學報、品管月刊等。
- d. 個人方面的論文著作。例如，博士論文、心得報告、臨床經驗等。

(2) 使用次級資料的優點：

- a. 既省時、省力又省錢。
- b. 可以獲得與研究問題有關的資料。
- c. 可以獲知其他相關資料的來源。
- d. 對所研究的問題有新的啟示。

(三) 資料的蒐集

1. 資料蒐集重點：統計工作所依據的全是統計資料，因此蒐集的資料是否客觀而周詳，關係整個統計工作的成敗。在蒐集統計資料時，須注意以下數點：確定調查的對象、時間及範圍、選定調查項目、實施方案、訓練與督導調查員等。例如，國家的人口普查，不適合在冬季的耶誕節與過年期間舉行。
2. 資料的蒐集來源：

- (1) 已存在資料（次級資料）。
- (2) 調查而得的資料（初級資料）。

(四) 資料的整理與表現

1. 資料整理：蒐集統計資料之後，應審慎而周詳地加以審核。如果發現資料中有錯誤、遺漏、殘缺、矛盾或其他可疑的地方，應迅速設法複查並加以更正。例如，蒐集來的資料單位不一致時，應加以換算，不能用的資料應捨棄；審核後的統計資料必須依照研究目的加以整理，以求簡化，並藉以顯示研究對象全體的通則。



2. 資料表現：為了便於說明，我們經常將資料經過整理分析後得到的統計結果以統計表或統計圖陳示，或另編統計報告詳加說明，以供各有關方面參考使用，如此統計工作程序即告完成。

四、資料的度量尺度

(一) 名目尺度

名目尺度指資料分類、屬性無大小之分。例如，性別、職業等。根據特徵的性質分類，每類答案的代表數字只作為分類用，目的在求分類。（此類資料可計算眾數）

(二) 順序尺度

順序尺度指資料有順序大小之分，但無差異程度。例如，名次、喜好程度等。在順序尺度中只能看出高低次序，但無法確定各類別之間的差距。（此類資料可計算中位數）

(三) 區間尺度

區間尺度指資料有大小、差異程度、無倍數關係。例如，溫度。兩個元素的資料值的比較，不僅可以表示順序的關係，也能測量各順序位置的距離。另外，區間尺度沒有自然原點 0。（此類資料可計算眾數、中位數、算術平均數）

(四) 比例尺度

比例尺度指資料有大小、差異和倍數關係。例如，身高、體重。比例尺度必包含零值。（此類資料可計算眾數、中位數、算術平均數）

※資料所含的資訊量因衡量尺度而異。資訊量由小排到大的順序是：名目 < 順序 < 區間 < 比例。區間與比例資料的算術運算有意義，名目與順序資料的算術運算沒有意義。

【觀念題 4】

試說明以下變數之測量方法屬於何種測量尺度？

- ①紅色 (0)，白色 (2)，藍色 (4)
- ②非常滿意 (2)，滿意 (1)，普通 (0)，不滿意 (-1)，非常不滿意 (-2)
- ③薪水（以元為單位）
- ④智商 (IQ 表示)

【解】

- ①名目尺度；②順序尺度；③比例尺度；④區間尺度。



五、定量資料的統計圖表

(一) 百分比、比例與比值

1. 百分比 (percentage)、比例 (proportion)：百分比與比例都是一種將資料標準化的作法。百分比是以 100 為基底來標準化，比例則是以 1.00 為基底。比例及百分比可用在名目尺度及順序測量的變項。

2. 比值 (ratio)：

$$\text{比值} = f_1 / f_2$$

f_1 是第一個類別或項目中之件數， f_2 是第二個類別或項目中之件數。

(二) 次數分配

1. 定義：一組資料之表列彙總，顯示在互不重疊的組別中所含物項的次數（或數目）。

2. 定量資料處理三步驟：

(1) 定組數：目標在於使用足夠的組數以顯示資料的變異性，但不希望組數過多致使每組僅有極少的元素，以 5 至 20 組為宜。

一般建議採用 Sturge's Rule，即組數 $k \approx 1 + 3.32 \log n$ 。

(2) 定組距：選定組數後，就可利用下列公式決定組距的近似值。

$$\text{組距近似值} = \frac{\text{最大的資料值} - \text{最小的資料值}}{\text{組數}}$$

(3) 定組限 (class limits)：下組限 (lower class limit) 是該組中可能的最小值，而上組限 (upper class limit) 則為該組中可能的最大值。

(三) 相對次數分配

相對次數分配指每組所含的資料次數（或數目）在全部資料次數 (n) 中所占的比率。

$$\text{相對次數} = \text{該組的次數} / n$$

【觀念題 5】

某飯店西餐廳 20 天晚餐結帳金額（單位：千元）如下：

32.65	26.12	42.90	34.26	42.00
24.10	29.86	39.40	38.75	72.00
28.40	34.50	69.80	64.45	61.81
36.62	46.12	53.00	53.06	49.42



試編製次數及相對次數分配表（分為 7 組）。

【解】

全距 = $72.00 - 24.10 = 47.90$ ，分為 7 組，所以取組距 = 7。

組別	組限	組界	組中點	劃記	次數	相對次數
1	24.00-30.99	23.995-30.995	27.495	\\ \\ \\	4	0.2
2	31.00-37.99	30.995-37.995	34.495	\\ \\ \\	4	0.2
3	38.00-44.99	37.995-44.995	41.495	\\ \\ \\	4	0.2
4	45.00-51.99	44.995-51.995	48.495	\\	2	0.1
5	52.00-58.99	51.995-58.995	55.495	\\	2	0.1
6	59.00-65.99	58.995-65.995	62.495	\\	2	0.1
7	66.00-72.99	65.995-72.995	69.495	\\	2	0.1
合計					20	1

(四) 直方圖

直方圖之長條間無空隙，可用於表示次數分配或相對次數分配。直方圖的橫軸是欲探究的變數，而縱軸是次數或相對次數。

1. 次數直方圖（Frequency histogram）：縱座標代表次數。
2. 相對次數直方圖（Relative frequency histogram）：以長方形（長條）的面積代表相對次數，因此長方形（長條）的高度 = 相對次數 $\div$ 組距，此高度通常稱為密度（density），縱座標尺度為密度尺度（density scale）。相對次數直方圖的所有長方形（長條）面積和為 1，亦即相對次數總和為 1。

【觀念題 6】

假設下列資料是截至 98 年 5 月底止本國 48 家銀行的分行數：

111	74	29	74	21	108	143	155
145	153	64	52	51	66	3	31
37	31	35	33	36	35	42	36
37	44	45	30	34	38	22	48
43	32	51	28	33	30	32	83
24	77	7	22	15	73	2	23



(1)作一個統計表，內容包括組別、組限、組中點、次數、相對次數、以下累加次數、以上累加次數，並說明編製步驟。

(2)繪製次數直方圖。

【解】

(1)統計表：

a. 求全距： $R = 155 - 2 = 153$ 。

b. 決定組數： $k = 1 + 3.32 \log 48 = 6.58 \approx 7$ 。

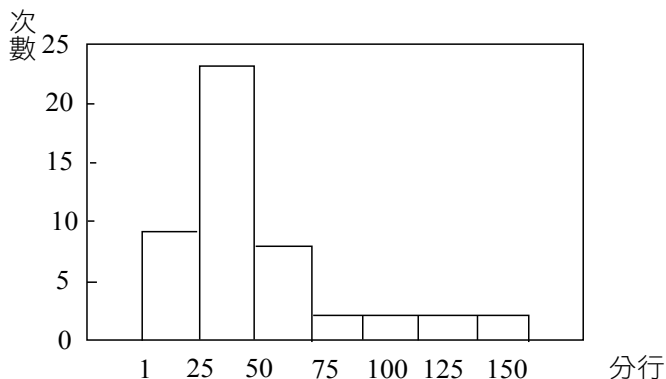
c. 決定組距： $\frac{153}{7} = 21.86$ ，取 25。

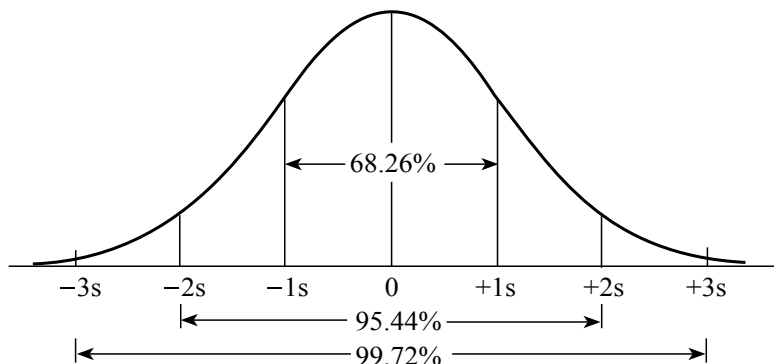
d. 決定組限，計算各組次數。

最小值為 2，選擇 1 為第一組組下限，25 為第一組組上限，以下類推，得次數分配表如下：

組別	組限	組中點	次數	相對次數	以下累加次數	以上累加次數
1	1-25	13	9	0.19	9	48
2	26-50	38	23	0.48	32	39
3	51-75	63	8	0.17	40	16
4	76-100	88	2	0.04	42	8
5	101-125	113	2	0.04	44	6
6	126-150	138	2	0.04	46	4
7	151-175	163	2	0.04	48	2

(2)次數直方圖：





私塾重點提示區

1. 統計圖表編製
2. 莖葉圖
3. 集中趨勢量數
4. 盒鬚圖／箱形圖

實力養成區

1. 請說明敘述統計學與推論統計學的差別，並分別舉一例說明之。

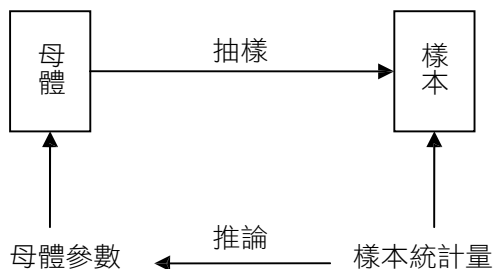
►►【解析】

- (1) 敘述統計學：包含資料的蒐集、整理、呈現、解釋和分析等步驟，一般以數值、表格、圖形來表現。例如，若想了解臺灣地區歷年來人口成長的情形，我們可以透過統計圖表將臺灣地區歷年來的人口數據資料表現出來。
- (2) 推論統計學：根據部分資料（稱為樣本）分析的結果對更大範圍資料（稱為母體）的某些特性，作一合理的推測和估計。例如，某次選舉民調得知，某候選人的支持率為 28%，則推論該地區民眾對此候選人的支持率為二成八。

2. 請說明母體參數與樣本統計量之間的關係，並以圖形說明之。

►►【解析】

當研究者欲估計或推論母體參數時，可先由母體中抽取適當之樣本，然後由所得之樣本計算其樣本統計量，最後再由樣本統計量來估計或推論母體參數，其關係可以下圖表示。



3. 下列資料表示一家雜貨店在過去兩個月中每日的顧客數：

24	59	80	33	56	75	51	71	81	37	79	78	80	96	40
66	91	86	56	75	53	10	65	74	79	24	81	97	82	66
42	70	84	55	63	73	87	63	73	44	59	54	88	77	83
49	83	91	16	78	35	66	18	81	70	73	64	79	86	60

請將此資料分成 9 組，最小一組由 10 開始。

試作一統計表，內容包括組別、組限、組界、次數、相對次數、以下累加次數、以上累加次數。

►►【解析】

組別	組限	組界	次數	相對次數	以下累加次數	以上累加次數
1	10-19	9.5-19.5	3	0.050	3	60
2	20-29	19.5-29.5	2	0.033	5	57
3	30-39	29.5-39.5	3	0.050	8	55
4	40-49	39.5-49.5	4	0.067	12	52
5	50-59	49.5-59.5	8	0.133	20	48
6	60-69	59.5-69.5	8	0.133	28	40
7	70-79	69.5-79.5	15	0.250	43	32
8	80-89	79.5-89.5	13	0.217	56	17
9	90-99	89.5-99.5	4	0.067	60	4

4. 一成衣製造公司的 25 名銷售員每週的銷售量紀錄如下所示，請繪出下列圖表。

24 56 43 35 37 27 29 44 34 28
33 28 46 31 38 41 48 38 27 29
37 33 31 40 50

- (1) 直方圖，使用 5 個類組。
- (2) 直方圖，使用 10 個類組。
- (3) 莖葉圖。



►►【解析】

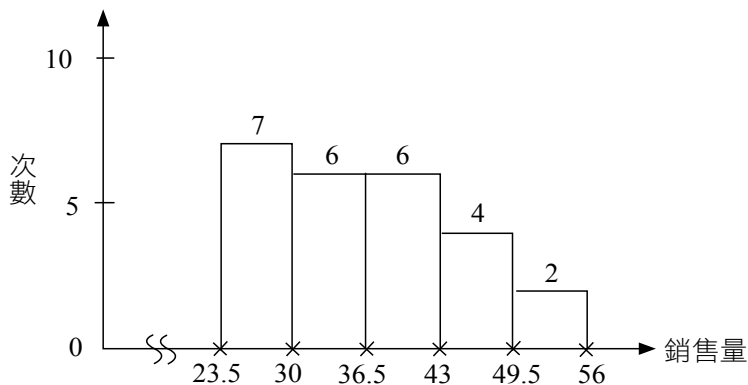
資料由小至大依序排列為：

24	27	27	28	28	29	29	31	31
33	33	34	35	37	37	38	38	40
41	43	44	46	48	50	56		

(1) 組距 $\frac{53-24}{5}=6.4$ ，取 6.5 之值，得次數分配表為：

銷售量組限	次數
23.5-30	7
30-36.5	6
36.5-43	6
43-49.5	4
49.5-56	2

直方圖為：



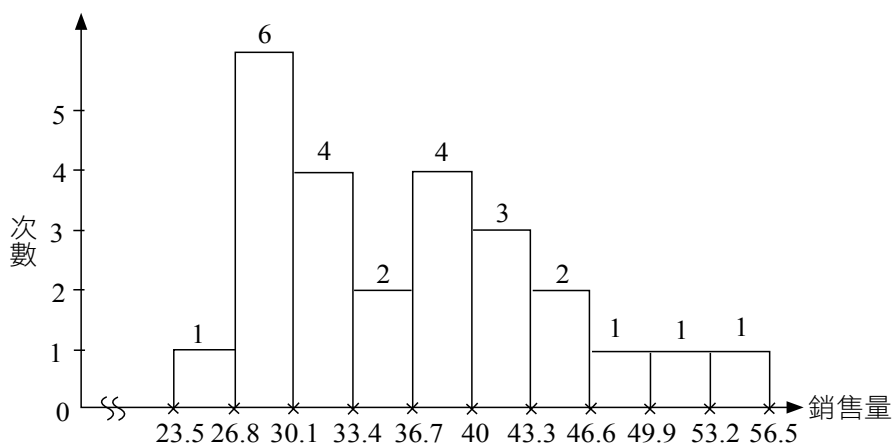
(2) 組距 $\frac{56-24}{10}=3.2$ ，取 3.3 之值，得次數分配表為：

銷售量組限	次數
23.5-26.8	1
26.8-30.1	6
30.1-33.4	4
33.4-36.7	2



36.7-40	4
40-43.3	3
43.3-46.6	2
46.6-49.9	1
49.9-53.2	1
53.2-56.5	1

直方圖為：



(3)

莖	葉
2	4798879
3	5743188731
4	346180
5	60

5. 某蠟燭生產商為測試其所生產的蠟燭燃燒時間，隨機選取 20 支蠟燭，其燃燒時間（單位：小時）結果如下：

4.10	4.09	3.98	4.26	4.14
4.25	4.24	4.08	4.25	4.28
3.97	4.18	4.22	4.25	4.26
4.00	4.10	4.24	4.27	4.36