



工程統計

李水彬Shuipin

2023/9/22

敘述統計(I)圖表法

學習重點

- 何謂分配? 知道資料的分配對決策有甚麼幫助?
- 資料有哪些類型? 衡量的尺度是甚麼?
- 如何展示資料的分配? 資料類型和衡量尺度對展示方法有何影響?

晶圓氧化層厚度

1516	1505	1448	1568	1428	1512	1601	1484	1483	1495
1512	1501	1462	1470	1427	1543	1470	1548	1547	1465
1472	1500	1482	1509	1425	1488	1459	1520	1379	1614
1449	1455	1412	1460	1517	1513	1451	1474	1461	1546
1533	1554	1500	1483	1460	1385	1569	1566	1469	1461

晶圓氧化層厚度

HTC	Samsung	Sony	HTC	Sony
iPhone	iPhone	iPhone	Huawei	HTC
HTC	Samsung	Nokia	HTC	Huawei
iPhone	Samsung	iPhone	iPhone	Samsung
Samsung	Sony	Sony	Blackberry	Samsung
HTC	Samsung	Asus	Blackberry	iPhone

資料的型態

- 人的屬性

姓名

生日

性別

血型

身高

體重

體溫

血壓(高)

血壓(低)

居住地

聯絡電話

學歷

職業

婚姻狀況

變數與變量

變數為可測量的個體特徵，變量為測量值。

- 我的 血型是 AB 型，的 身高 175 公分。
- 今天 就診人數為 79 人。
- 我的 手機品牌為 xCall, 他的 手機品牌為 yNet。

變數的屬性

- **屬質變數**: 與他人不同的特性, 以作為區分之用。其目的僅於分類, 故又稱為 **類別資料** (categorical data) 。
 - 姓名, 性別, 血型, 居住地, 學歷, 職業, 學歷, 和婚姻狀態這些變數都屬此類。
- **屬量變數**: 『量』有 **數量** 的意思, 數量可增減, 有多有少。量的資料為衡量個體在一個變數上擁有的多寡、大小。
 - 身高、體重、體溫、血壓等皆為屬量變數。

屬質變數

- 血型有 O, A, B 和 AB 型之別。
- 婚姻狀況分成已婚與未婚之別。
- 學歷有小學, 國中, 高中, 大學, 研究所以上之別。
- 若一生產線有很多同機台, 給予機台編碼區分, 如機台 A, 機台等。

屬量變數

- **離散型變數**: 計數型資料, 紀錄某一種特性發生的次數。例如, 班上人數, 缺曠人數, 不良品個數, 某日車禍發生次數, 死亡人數, 新生兒人數等等。
- **連續型變數**: 非計數型資料的屬量變數為連續型變數, 連續者, 接續不斷之意, 是一個極為抽象的概念。量測實體 **不能被切割成一些最小單位的組成**。例如, 身高、體重等。

L11的宅記-1

昨天,領到打工的薪水 23023 元,立馬買一整套炫彩黃色籃球背心與短褲,配上我的Nike 籃球鞋,整個人帥到爆,昨晚約阿丹,佐佐和可樂要來學校練球。

今天天氣超熱了,早上10點溫度計顯示 32 度C。根據 BBC 報導,在這個溫度下打網球每小時會爆汗 3 公升,我想打籃球可能會更多,

所以,我準備 5 罐大罐的運動飲料共6250cc 來到健行科大的操場。可是,(時間)已經過了 10 分鐘操場一個人都沒,拿起手機一看,佐佐 LINE 我說,他們要跟 2 位女同學小惠,君宜去游泳池避暑了。

L11的宅記-2

早上女友說我好像胖了許多,應該要做運動減肥少吃吃一點。哇,體重計是不是壞了,82.3公斤太可怕了。

Google一下,籃球每小時每公斤6.3大卡,慢跑每小時每公斤8.2大卡,所以,我打1小時籃球可以消耗518.5大卡,慢跑1小時可以消耗674.9大卡。

Go,我們運動去。可是天氣實在太熱,還是去吃紅豆牛奶冰吧,運動就交給明天負責。

分不清是連續還是離散

屬量變數又可分為離散型和連續型變數兩種。我們不應從資料出現表徵, 作為辨識變數是否連續的依據, 而它往往只是一種 [測量上的限制](#), 反應量測系統的能力。

身高 175 分和 175.3 公分, 我們不能因為前者有沒有小數點, 就說身高是一個離散型變數, 後者有小數一位就說身高一個連續型變數。身高是一個長度單位, 紀錄的有效位數是 [量測系統精確度](#) 的問題, 不是變數本質的問題。長度本質是連續的, 所以身高就是連續變數。同理, 體重, 溫度, 時間, 血壓等變數都是連續型變數。

連續或離散不是很重要

考試分數是一個很奇怪的變數, 分數往往是一分一分往上累加, 但有時最小單位不一定是一分, 考題是選擇題或填充題, 考試分數對應答對的題數, 它是一個離散變數。

例如老師考50題的選擇題, 每題2分, 學生成績的最小單位就是兩分。但有時考問答題或計算題, 老師的給分會出現小數, 而且沒有一個明確的最小單位, 在統計分析時, 可以將考試得分看成連續變數。一筆數據的複雜度決定將變數看成離散或連續, 可能情況越多表示越複雜。

例如, 一個班級考試分數有60, 70, 80, 90四種, 另一個班級考試分數有 40, 42, 43, 50, 51, 53, 66, 68, 70, 75, 80, 92。把後者看成連續是可以的。

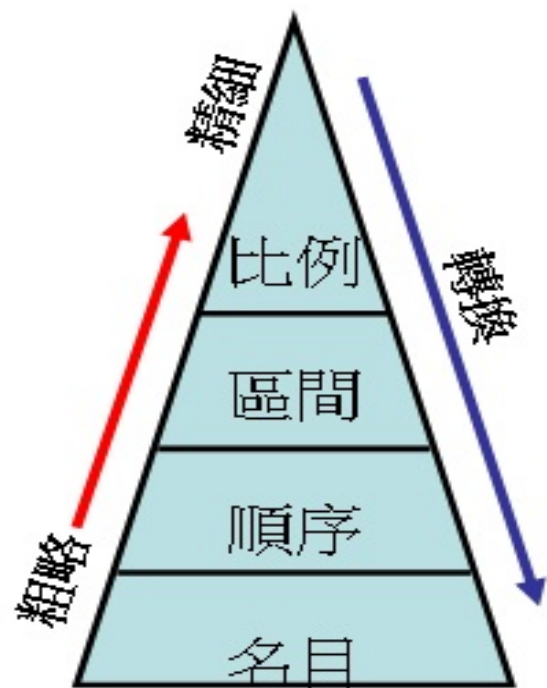
衡量的尺度

- **名目尺度:** 區分不同,將群體分成數個互斥子群體的衡量方式,即稱為名目尺度。
- **順序尺度:** 呈現程度上的差異。
- **區間尺度:** 把差異的程度數量化,數值為相對數量。
- **比例尺度:** 數值為絕對數量。

常用變數的衡量尺度

- **名目尺度:** 姓名、血型、宗教信仰、膚色、婚姻狀況、職業、手機品牌
- **順序尺度:** 滿意度、(考試成績)排名、教育程度、蒲福風級數, 等級
- **區間尺度:** 溫度、西元年。
- **比例尺度:** 身高、體重、時間、壓力、次數、個數、人數、缺點數、不良數、通話時間

尺度轉換



衡量尺度與四則運算的關係






- 名目和順序尺度沒有量的意義, 故在資料處理上, 不能使用加減乘除四則運算。通常合適的做法, 僅計算每一個類別觀察到的次數。
- 區間尺度沒有絕對的原點, 故無法以『比』值描述數字間的關係, 換言之, 不能做『乘除』的運算。例如, 我們說今天的氣溫上升 5 度, 但不能說氣溫比昨天高出 2 倍。

次數分配表



Watermelon	Apple	Watermelon	Grape	Banana	Apple	Strawberry
Apple	Grape	Watermelon	Banana	Apple	Watermelon	Watermelon
Banana	Watermelon	Strawberry	Watermelon	Apple	Strawberry	Banana
Banana	Strawberry	Watermelon	Banana	Watermelon	Grape	Grape
Strawberry	Apple	Banana	Apple	Banana	Apple	Watermelon

掛環種類的次數分配表

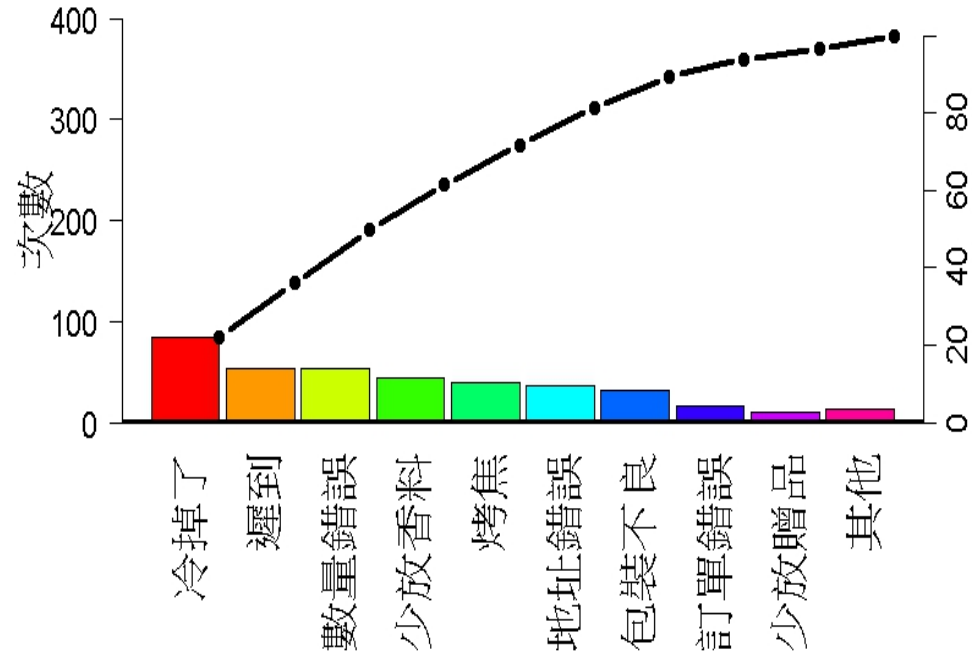
吊飾種類						合計
個數	10	8	8	5	4	35
相對次數	0.29	0.23	0.23	0.14	0.11	1.0

顧客抱怨(屬質變數)的次數分配表

- 統計每種抱怨發生的次數

冷掉了	地址錯誤	遲到	包裝不良	數量錯誤	訂單錯誤	少放香料	少放贈品	烤焦	其他
85	37	53	31	53	17	44	10	40	13

柏拉圖



何時使用柏拉圖

- 柏拉圖是 **品管七大手法**之一, 旨在找出主要問題。

繪製柏拉圖的程序

- 次數分配表
- 依發生頻率排序
- 將 其他 排到最後
- 計算累積次數
- 計算相對次數和相對累積次數
- 繪製次數分配的長條圖，y軸座標在左邊
- 加入累積次數的折線圖，y軸座標在右邊

顧客抱怨的次數分配表

冷掉了	地址錯誤	遲到	包裝不良	數量錯誤	訂單錯誤	少放香料	少放贈品	烤焦	其他
85	37	53	31	53	17	44	10	40	13

依發生頻率排序

冷掉了	遲到	數量錯誤	少放香料	烤焦	地址錯誤	包裝不良	訂單錯誤	其他	少放贈品
85	53	53	44	40	37	31	17	13	10

將 其他 排到最後

冷掉了	遲到	數量錯誤	少放香料	烤焦	地址錯誤	包裝不良	訂單錯誤	少放贈品	其他
85	53	53	44	40	37	31	17	10	13

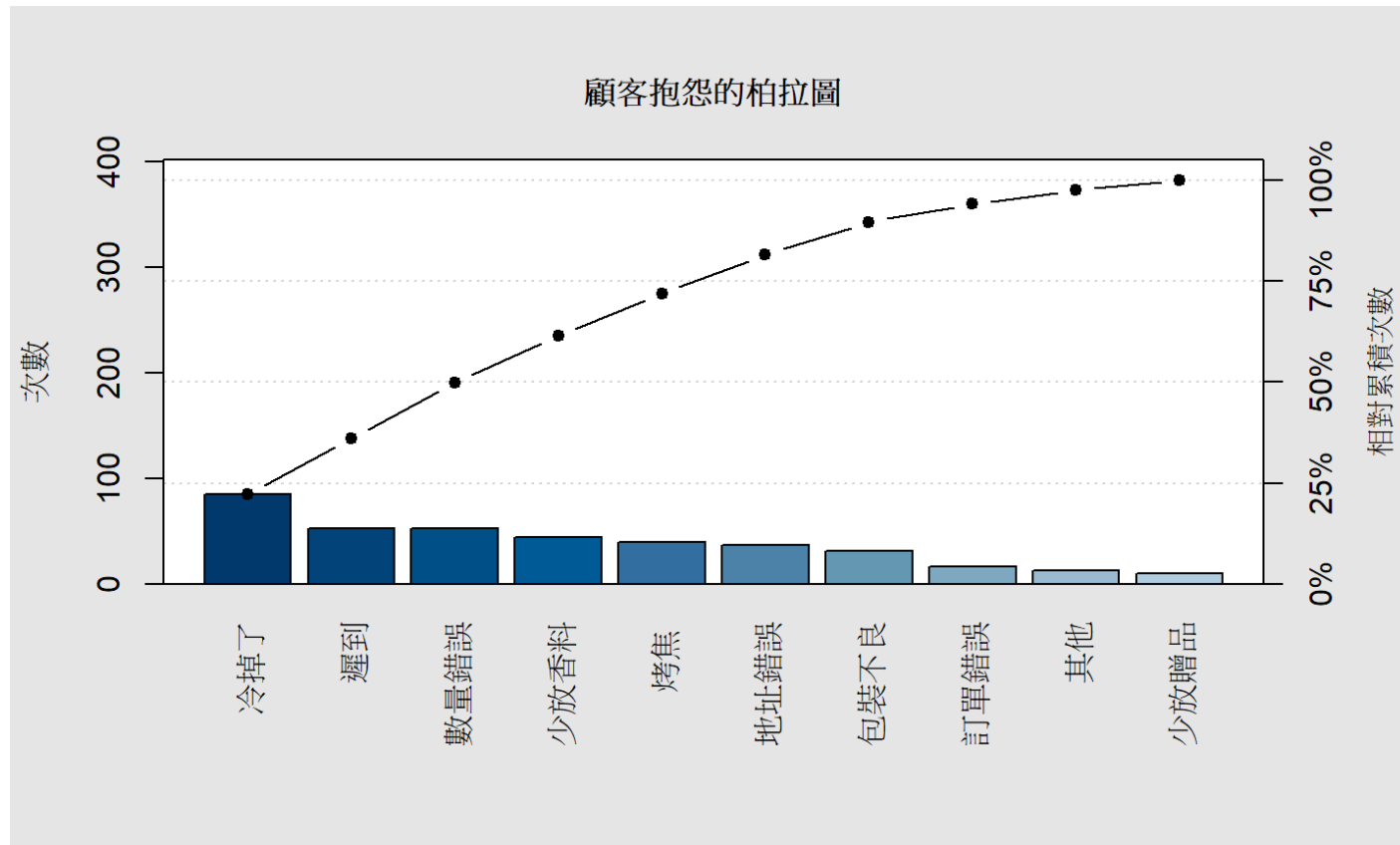
計算累積次數

	冷掉了	遲到	數量錯誤	少放香料	烤焦	地址錯誤	包裝不良	訂單錯誤	少放贈品	其他
次數	85	53	53	44	40	37	31	17	10	13
累積次數	85	138	191	235	275	312	343	360	370	383

計算相對次數與相對累積次數

	冷掉了	遲到	數量錯誤	少放香料	烤焦	地址錯誤	包裝不良	訂單錯誤	少放贈品	其他
次數	85.0	53.0	53.0	44.0	40.0	37.0	31.0	17.0	10.0	13.0
累積次數	85.0	138.0	191.0	235.0	275.0	312.0	343.0	360.0	370.0	383.0
相對次數	22.2	13.8	13.8	11.5	10.4	9.7	8.1	4.4	2.6	3.4
相對累積次數	22.2	36.0	49.9	61.4	71.8	81.5	89.6	94.0	96.6	100.0

柏拉圖



```
##  
## Pareto chart analysis for dd[, 1]  
##           Frequency  Cum.Freq.  Percentage  Cum.Percent.  
## 冷掉了      85.000000  85.000000   22.193211   22.193211  
## 遲到        53.000000  138.000000  13.838120   36.031332  
## 數量錯誤    53.000000  191.000000  13.838120   49.869452  
## 少放香料    44.000000  235.000000  11.488251   61.357702
```

屬量資料的整理

- 離散型 (計數型): 統計每一種情況發生的次數和頻率。

6	1	7	3	1	2	3	1	5	3
1	2	2	1	3	2	5	4	6	2
1	2	0	2	1	3	2	3	2	2
6	2	2	5	1	4	5	3	3	1
2	2	3	0	3	4	5	5	3	2

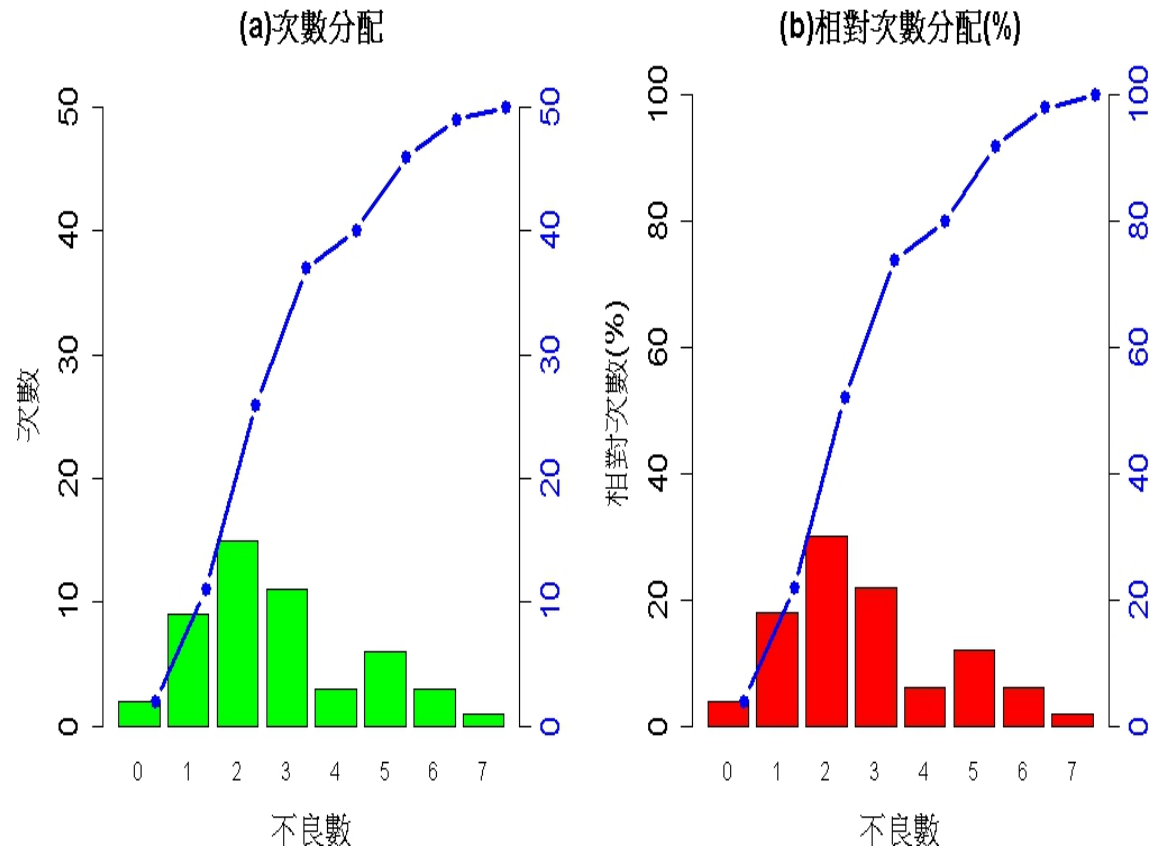
次數分配表

不良數	畫記	次數	累積次數
0		2	2
1		9	11
2		15	26
3		11	37
4		3	40
5		6	46
6		3	49
7		1	50

相對次數分配

不良數	畫記	次數	累積次數	相對次數	累積相對次數(%)
0		2	2	0.04	4
1		9	11	0.18	22
2		15	26	0.3	52
3		11	37	0.22	74
4		3	40	0.06	80
5		6	46	0.12	92
6		3	49	0.06	98
7		1	50	0.02	100

次數分配圖



連續型變數- 直方圖

直方圖是用來展現連續型變數的數據分配, 製作程序是先完成次數分配表後再繪製直方圖。

製作次數分配表

1. 找出最大數據 (Max) 與最小數據 (Min) 來。
2. 計算數據的全距 $R = Max - Min$ 。
3. 決定適當組數 k ($5 \sim 20$), 與組距 h 。注意: $k \times h > R$ 。
4. 求出最小組的下組界 $s_1 \leq Min$, 即最小組的下組界必須小於或等於 Min 。
5. 算出每一組之下組界 $s_i = s_{i-1} + h$ 與上組界 $l_i = s_i + h, i = 1, 2, \dots, k$ 。檢查是否最大值是否落在最後一組內?如果不是, 必須調整最小組的下組界或組距。
6. 計算各組的組中點 $m_i = s_i + h/2, i = 1, 2, \dots, k$ 。
7. 將數據畫記於表內。
8. 合計次數是否正確?