

資料敘述級

連續資料

中心與離度

巨量資料探勘與統計應用 W05

資訊的中心與離度

布丁布丁吃布丁  
<http://blog.pulipuli.info/>



# 最適合你的一個字



# 最適合你的一個字

布丁

2017年最適合你的一個字

狃

你的做事效率機會大幅增加  
生活的轉折也會有驚人的變化

將會是讓別人對你完全  
耳目一新的一年



汽車與卡車...

pulipuli.chen@gmail.com

檔案 編輯 檢視 插入 格式 註解 共用 說明

所有變更都已儲存到雲端硬碟

fx | 汽車價格(千美元)

	A	B	C	D
1	汽車價格(千美元)	卡車價格(千美元)		
2	22	46		
3	28	19		
4	42	21		
5	24	19		
6	34	17		
7	62	26		
8	27	20		
9	33	32		
10	39	21		
11	22	36		
12	25	12		
13	32	27		
14	28	5		
15	40			
16	44			
17	40			

data



# 汽車跟卡車的價格

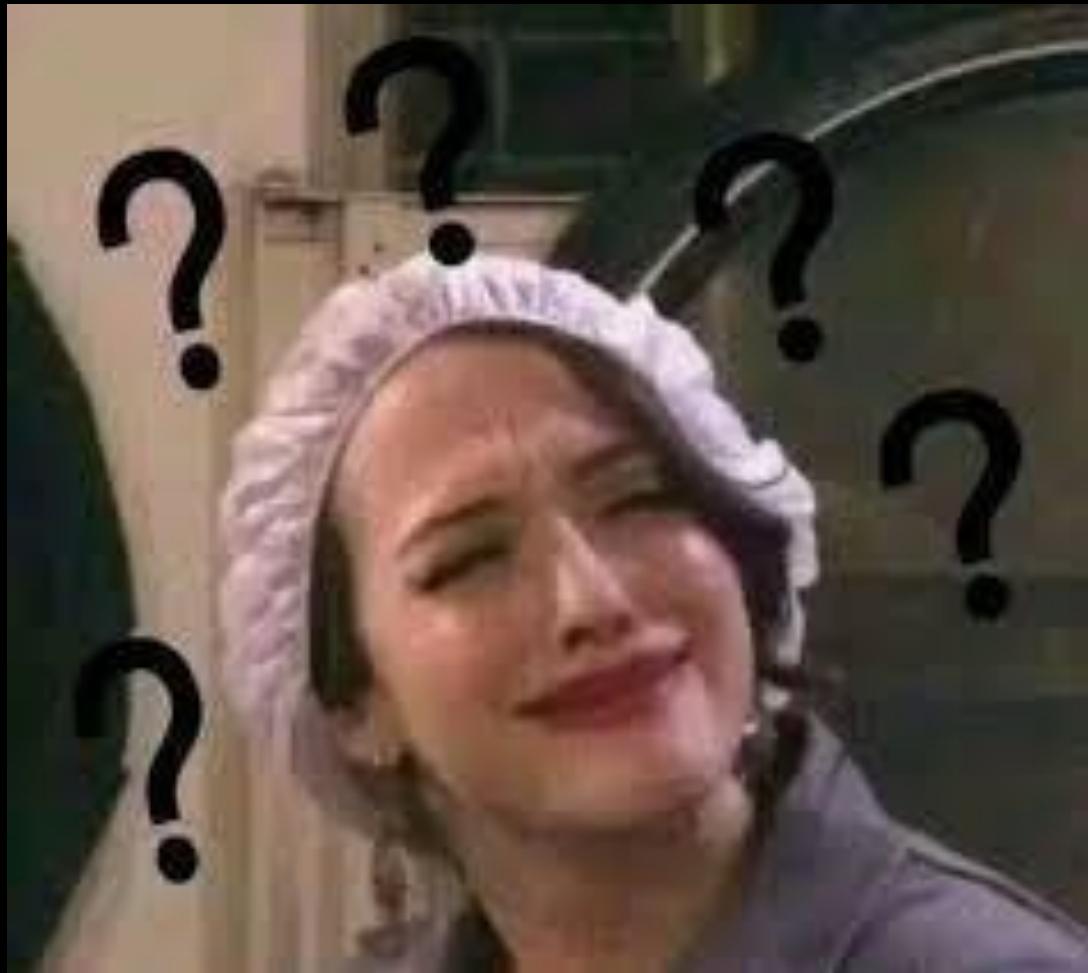
汽車案例數量:100筆

卡車案例筆數:40筆

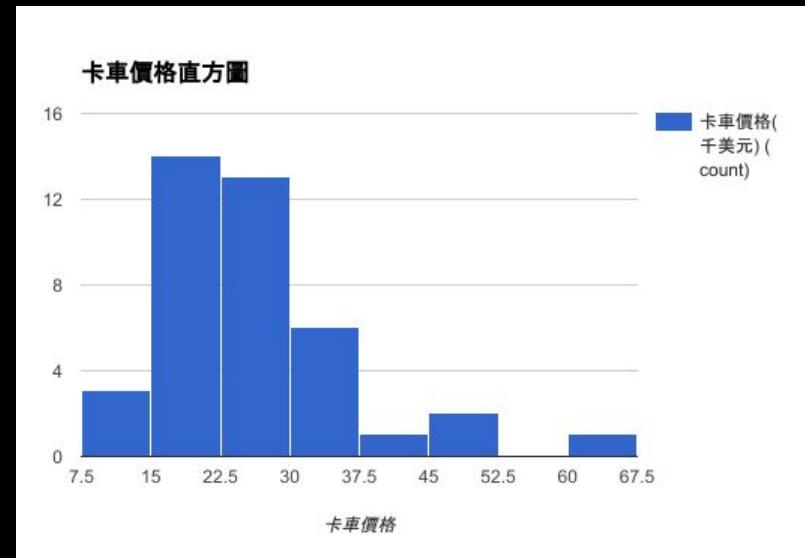
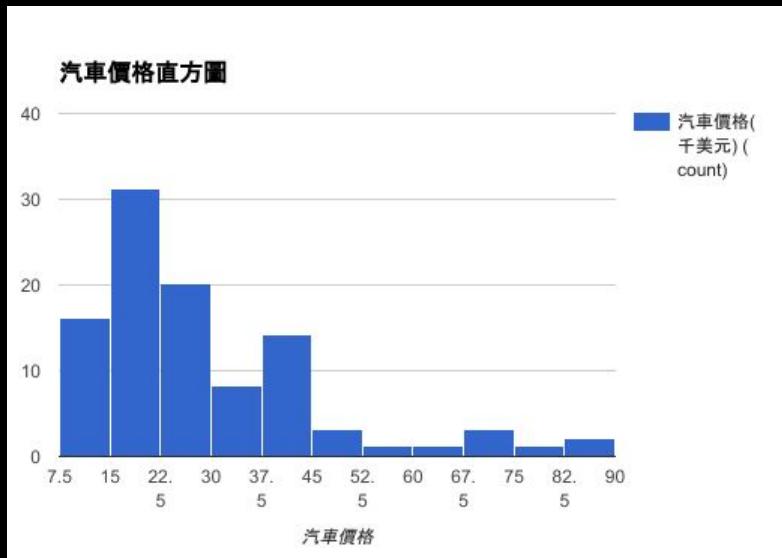
# 本週的老闆



「我不想看圖，你給我用講的介紹這些資料。」



統計圖表不好嗎？



這要怎麼比較啊？

# 資料敘述級分析的步驟

1. 把資料畫成圖
2. 加入代表性的綜合數值
3. 尋找資料的整體形態、以及有異與整體形態的偏差
4. 當資料整體形態很有規律時，用精簡的方式來描述它

# 資料科學家的技能之一



看穿資料的綜合戰鬥力數值

中心 & 離度

# 本週課程大綱

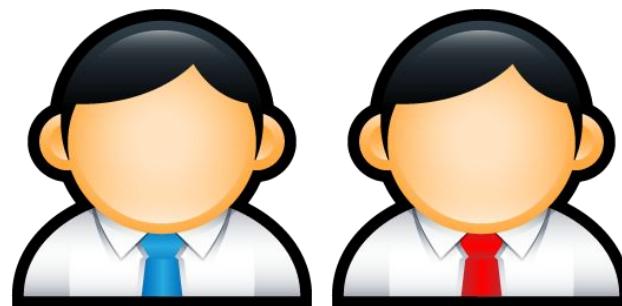
1. 資料的中心:均值
2. 實作:均值的計算
3. 資料的離度:分散性
4. 實作:五數綜合的計算
5. 實作:五數綜合的繪圖
6. 資料的離度:變動性
7. 實作:變異數與標準差的計算
8. 練習:不同年齡對網路郵局的滿意度

Part 1.

# 資料的中心:均 值

# 黑心企業？

週薪均值  
**2500美元**  
我們要加薪！



員工  
\$500

員工  
\$500

# 黑心企業？

週薪均值

**14000美元**

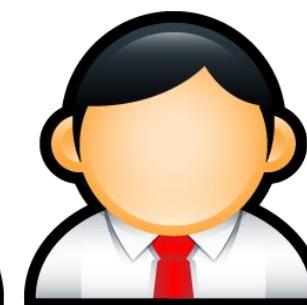
我也要加薪！



經理  
\$4500

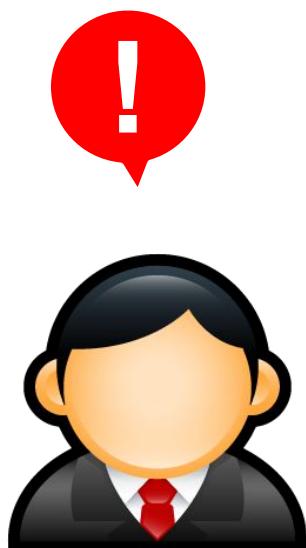


員工  
\$500

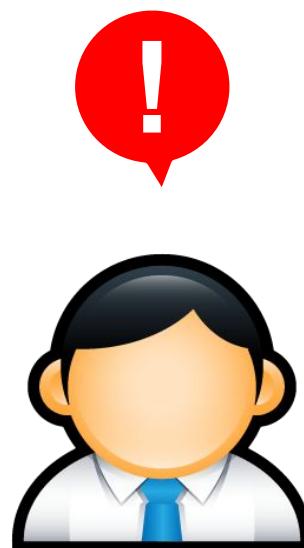


員工  
\$500

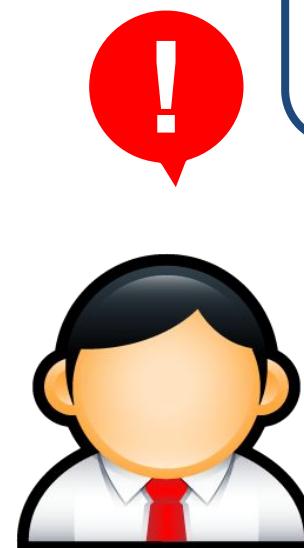
# 黑心企業？



經理  
\$4500



員工  
\$500



員工  
\$500



CEO  
\$50500

週薪均值  
**500美元**而已！  
回去工作！

# 誰在說謊？



# 誰在說謊？

平均數

週薪均值  
**14000美元**



經理  
\$4500

中位數

週薪均值  
**2500美元**



員工  
\$500



員工  
\$500

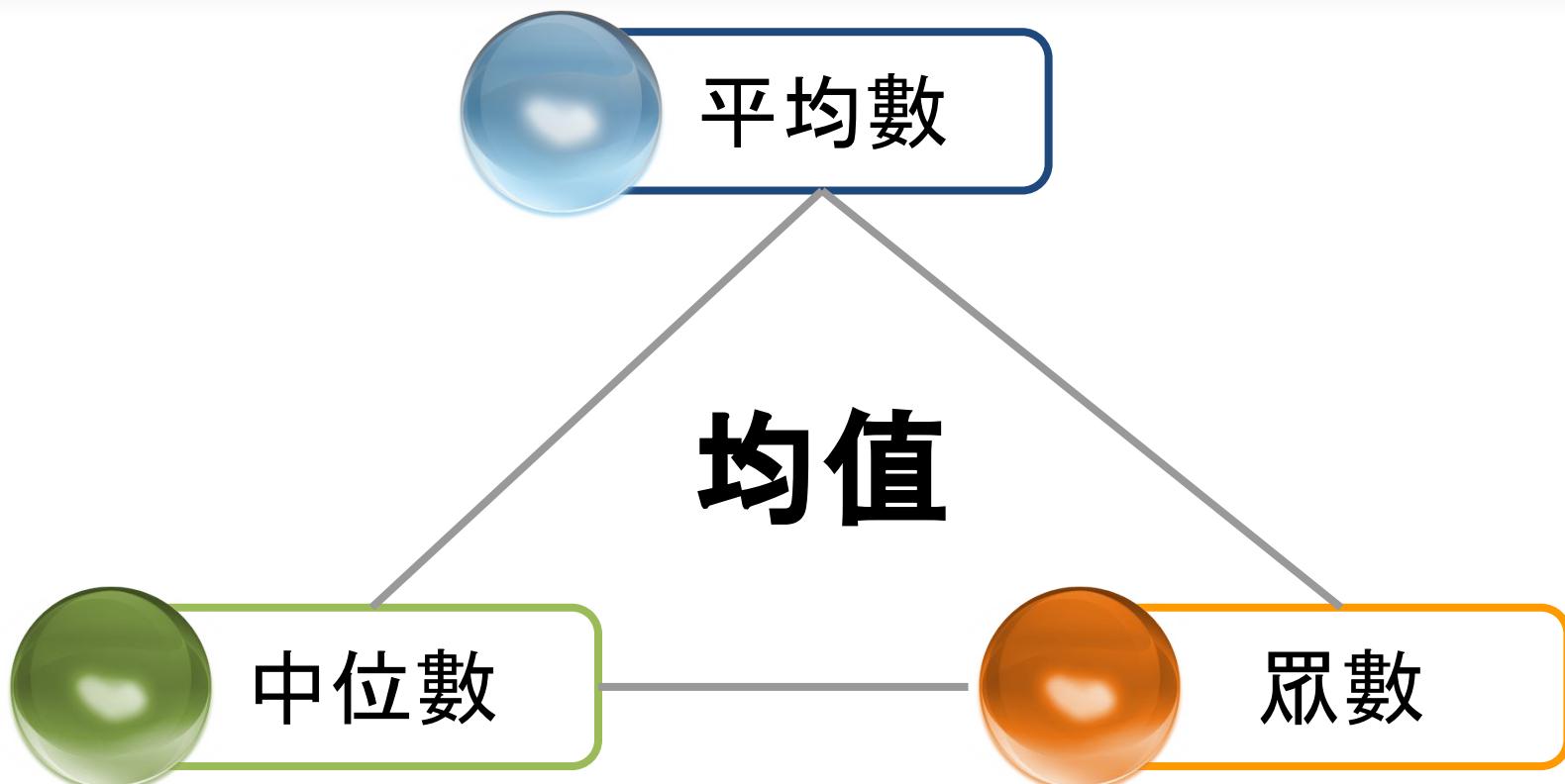
眾數

週薪均值  
**500美元**

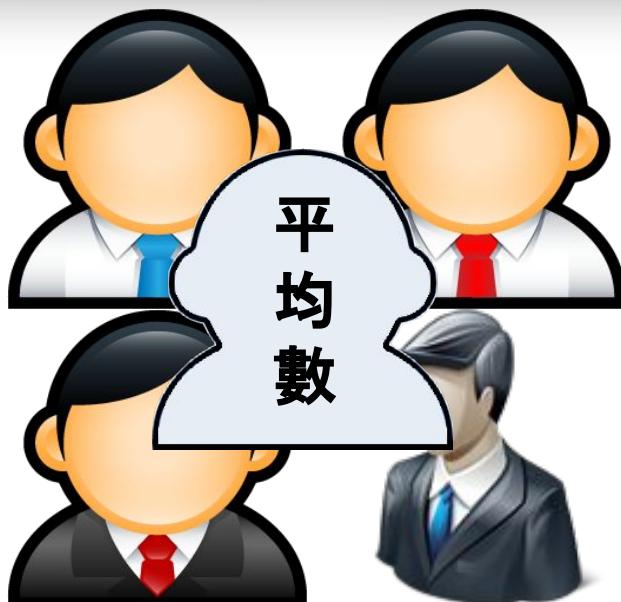


CEO  
\$50500

# 資料中心的三神器 平均數, 中位數, 眾數



# 平均數



- 算數平均數  $\bar{x}$  ( $x$ -bar)
  - 連續資料加總
  - 除以資料筆數  $n$
- 平均數是一種抽象化的數值

週薪平均數 =

$$(500+500+4500+50500)/4\text{人} =$$

**14000**

$$\bar{x} = \frac{n \text{ 個觀測值的和}}{n}$$



=AVERAGE()

# 平均數常用的場合

## 民國104年 平均每戶家庭人數

家庭戶數	8,386,495
平均每戶人數	3.1
平均每戶成年人數	2.53
平均每戶就業人數	1.43



## Barry Lamar Bonds 在大聯盟的全壘打數

16	25	24	19	33
25	34	46	37	33
42	40	37	34	49
73	46	45	45	5
26	28			

平均數=34.6

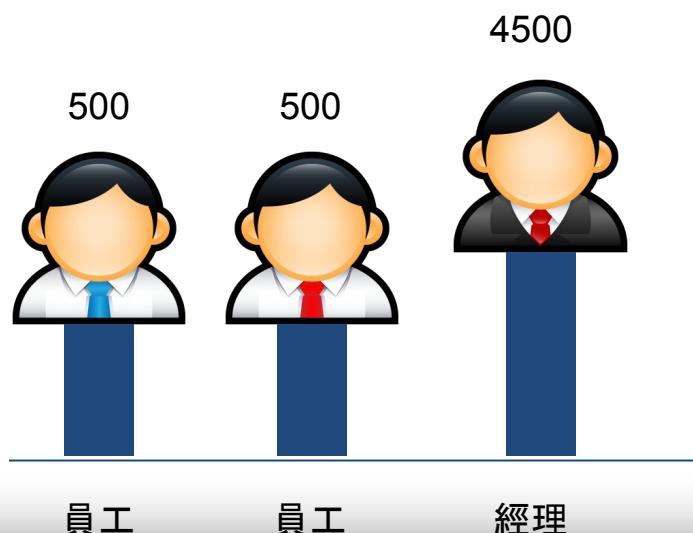


# 平均數的問題

離群值

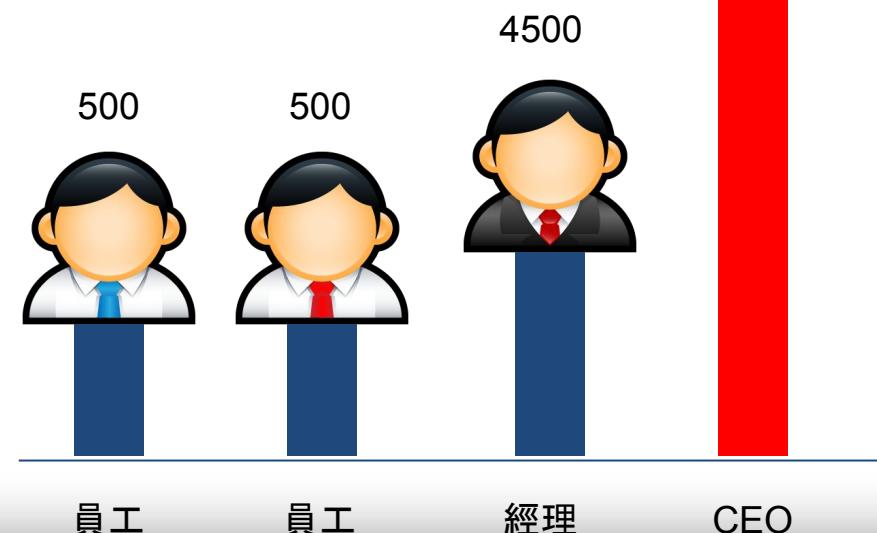
平均數:

1833.33



平均數:

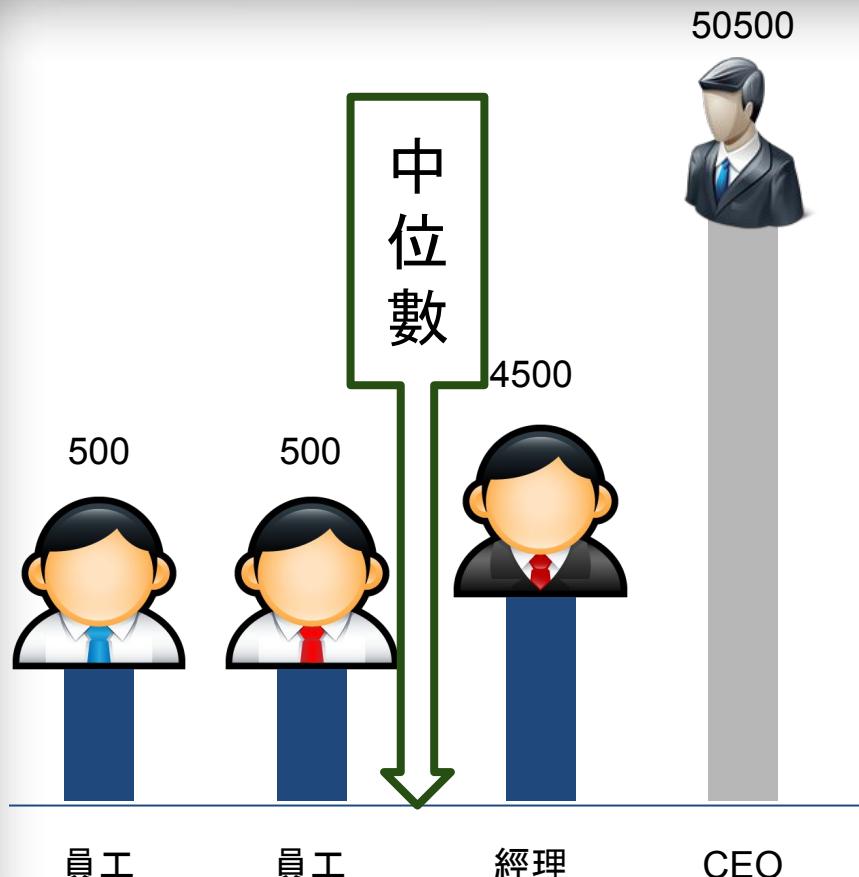
14000



高薪錯了嗎？



# 中位數



- 中位數是資料分佈的中間點
  - 一半資料比它小
  - 一半資料比它大

週薪中位數 =  
排行第二 500 & 排行第三 4500  
的中間數值 = **2500**

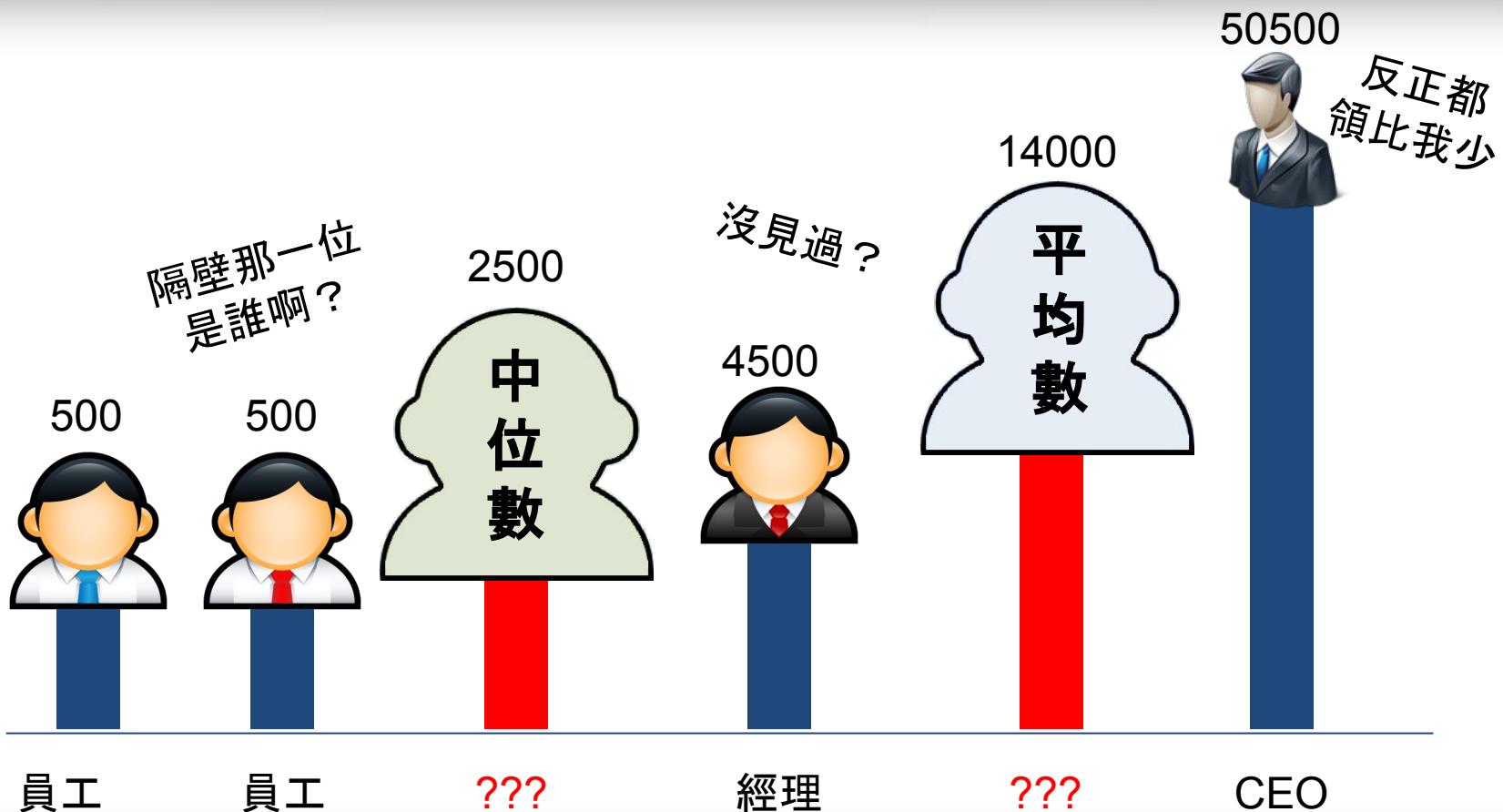


# 中位數的使用場合

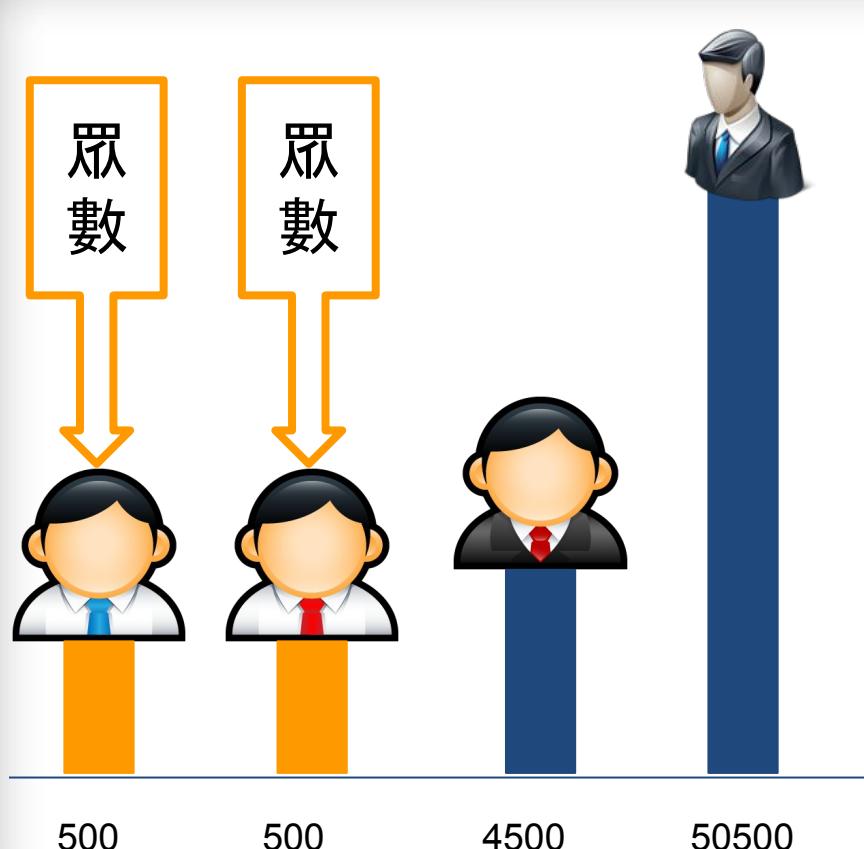
工業及服務業受僱員工每人每月薪資及經常性薪資  
(單位:新臺幣元)

	薪資(含經常性與非經常性薪資)				經常性薪資			
	服務業部門		工業部門		服務業部門		工業部門	
	中位數	年增率(%)	中位數	年增率(%)	中位數	年增率(%)	中位數	年增率(%)
99年	37 756	4.43	44 359	5.16	31 664	1.78	36 214	1.64
100年	38 870	2.95	45 508	2.59	31 505	-0.50	36 689	1.31
101年	38 632	-0.61	45 589	0.18	31 800	0.94	37 151	1.26
102年	38 625	-0.02	45 664	0.16	32 132	1.04	37 527	1.01
103年	40 007	3.58	47 300	3.58	32 702	1.77	38 208	1.81
104年	40 853	2.11	48 490	2.52	33 434	2.24	38 716	1.33

# 平均數跟中位數的限制



# 眾數

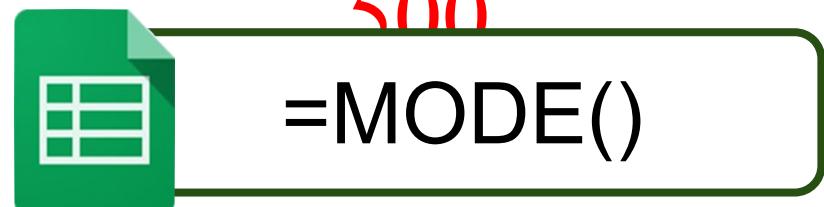


眾數是指最普遍存在的值

- 曊數一定是實際存在的數值
- 曊數可能不只一個
- 曊數可以描述類別資料

週薪眾數 =

500 出現最多次(2次) =



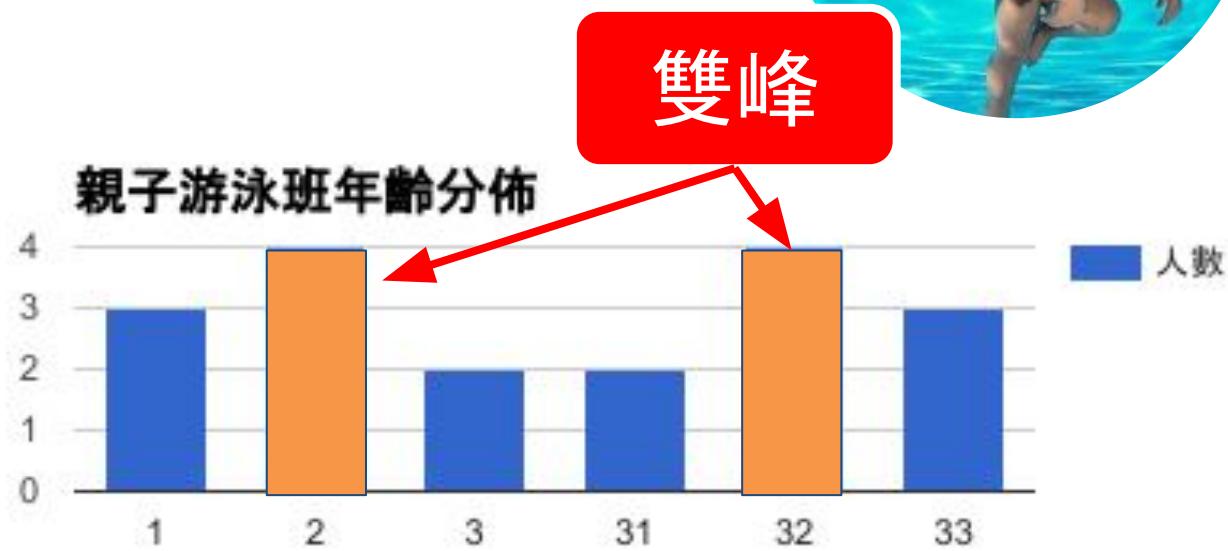
# 眾數的特性：可以有多個眾數

能代表這個親子游泳班年齡的有

2歲 & 32歲



雙峰



# 眾數的特性：可以處理類別資料

圖選欄			
號次欄	1	2	3
相片欄	小丑糖	蜂糖	幼兒糖
姓名欄	犯群喵黨	魅力無法黨	八點黨
推薦政黨欄			

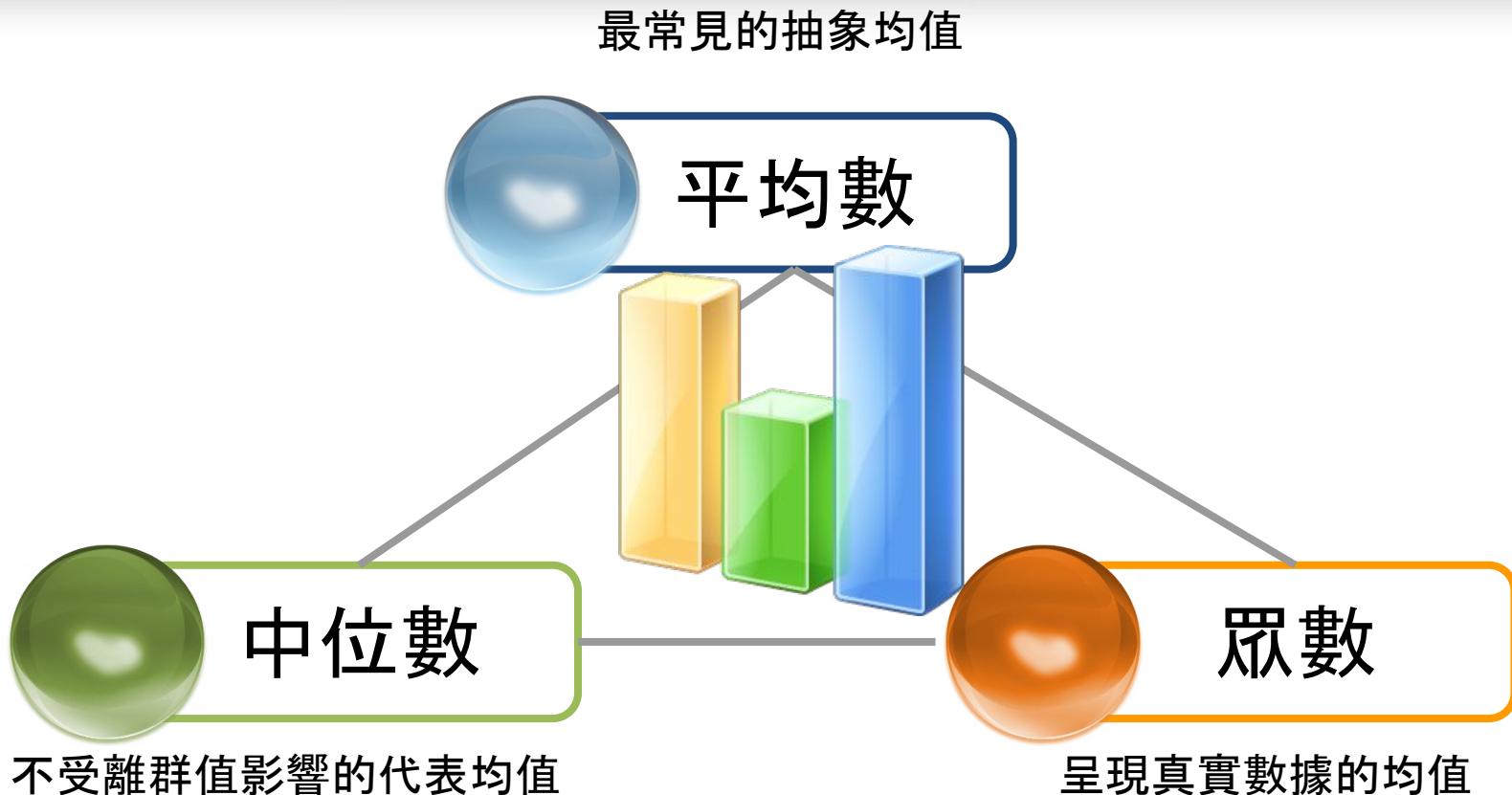


小丑糖
蜂糖
幼兒糖
蜂糖
幼兒糖
蜂糖
蜂糖
小丑糖
幼兒糖
蜂糖

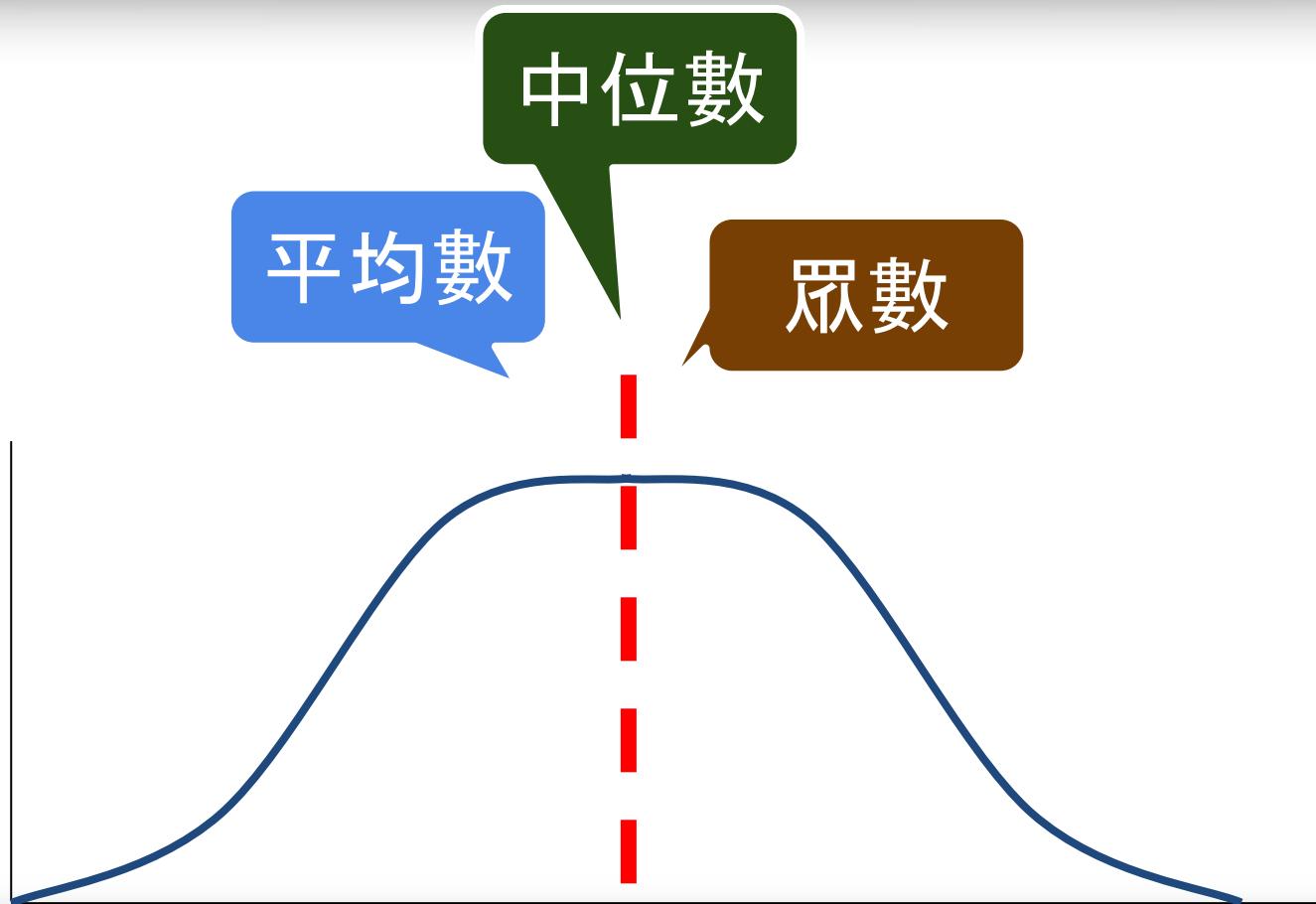


眾數  
蜂糖

# 均值與資料形狀的關係



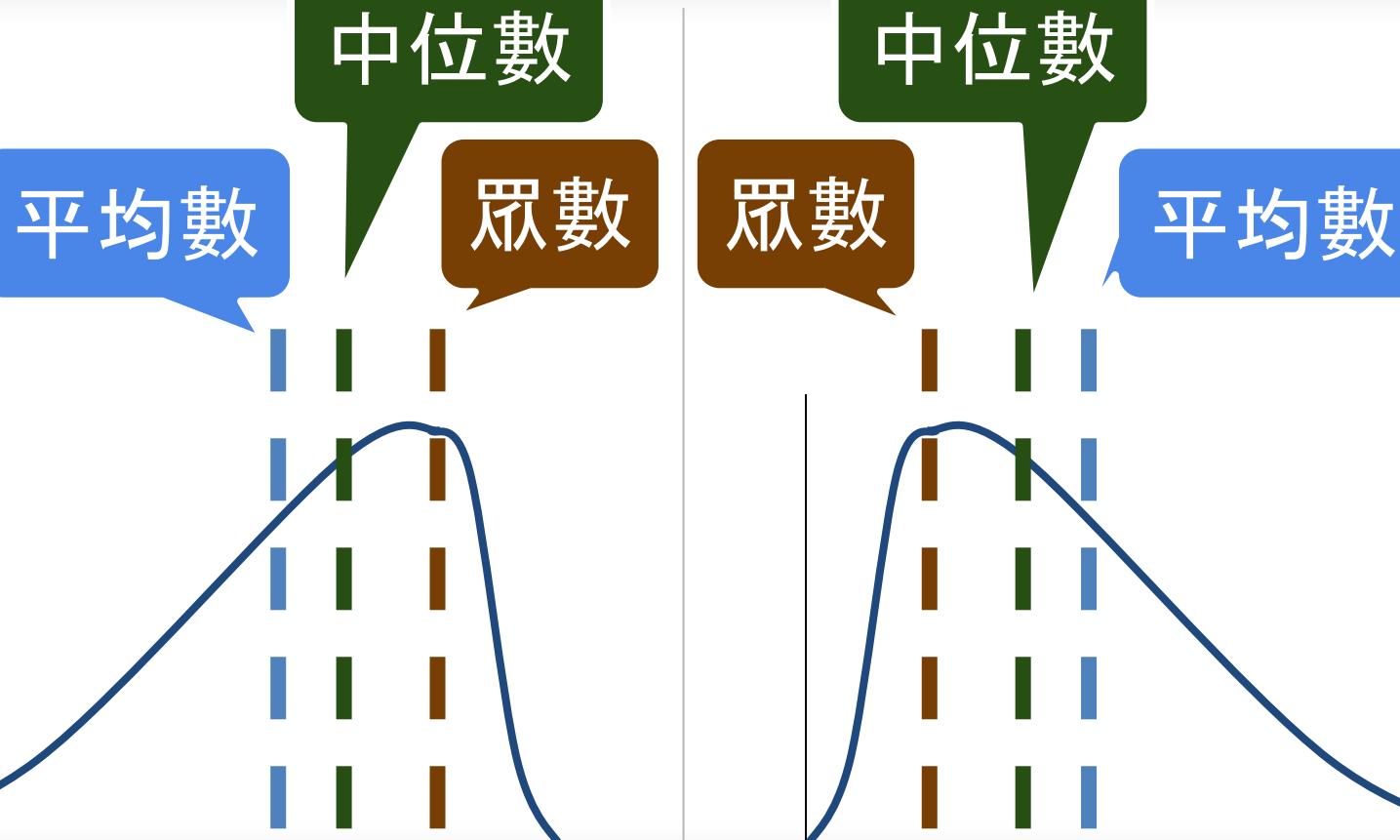
# 均值與資料形狀的關係 對稱=常態分配



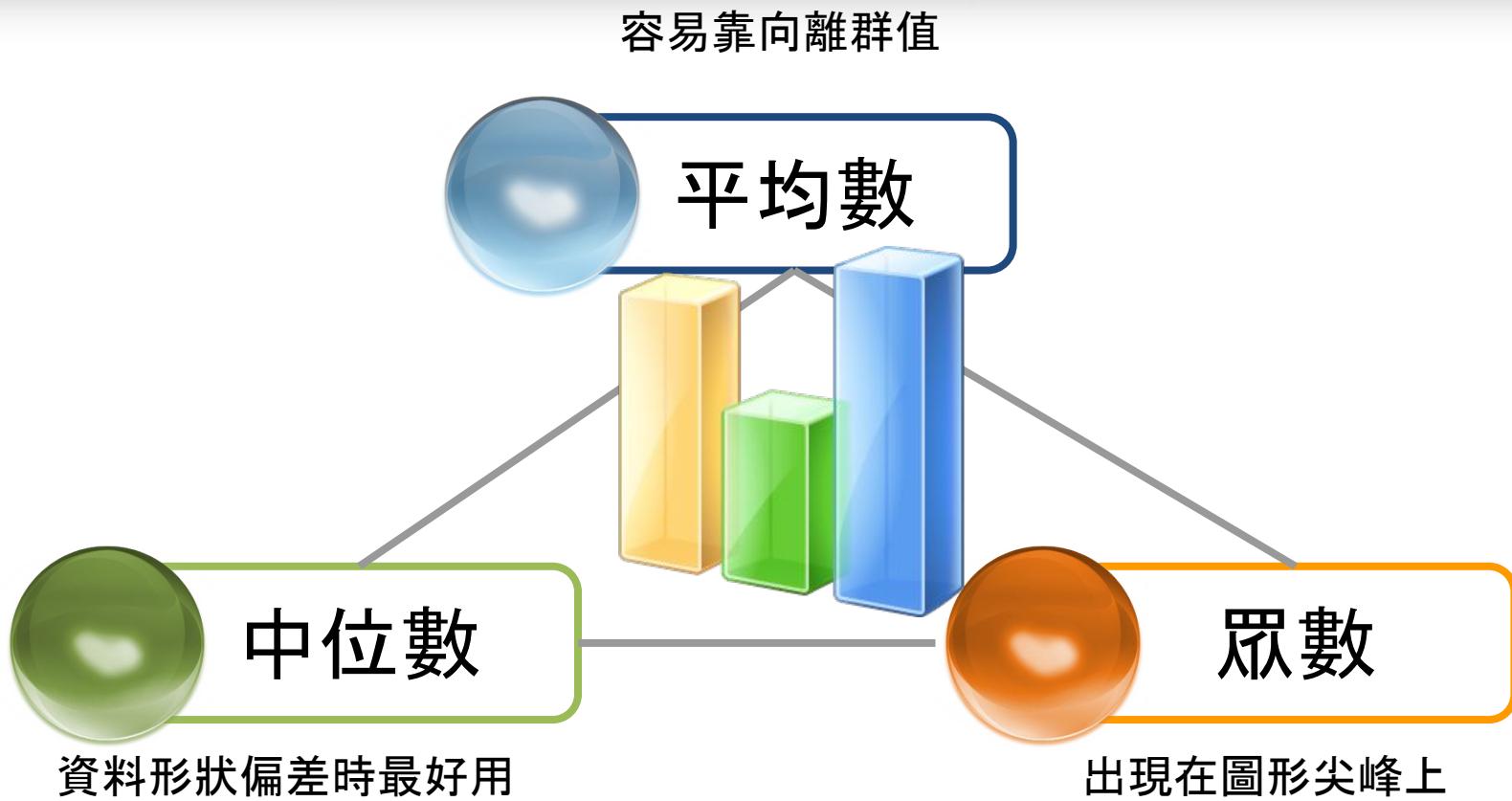
# 均值與資料形狀的關係

左偏 (成績)

右偏 (收入)



# 均值與資料形狀的關係



Part 2.

## 實作:均值的計算

# 聽說你學了新招式



「來幫我算算汽車價格的代表綜合數 值吧。」

汽車與卡車...

pulipuli.chen@gmail.com

檔案 編輯 檢視 插入 格式 註解 共用 說明

所有變更都已儲存到雲端硬碟

fx 汽車價格(千美元)

	A	B	C	D
1	汽車價格(千美元)	卡車價格(千美元)		
2	22	46		
3	28	19		
4	42	21		
5	24	19		
6	34	17		
7	62	26		
8	27	20		
9	33	32		
10	39	2		
11	22			
12	25			
13	32			
14	28			
15	40			
16	44			
17	40			

data



# 汽車的價格

汽車案例數量: 100筆

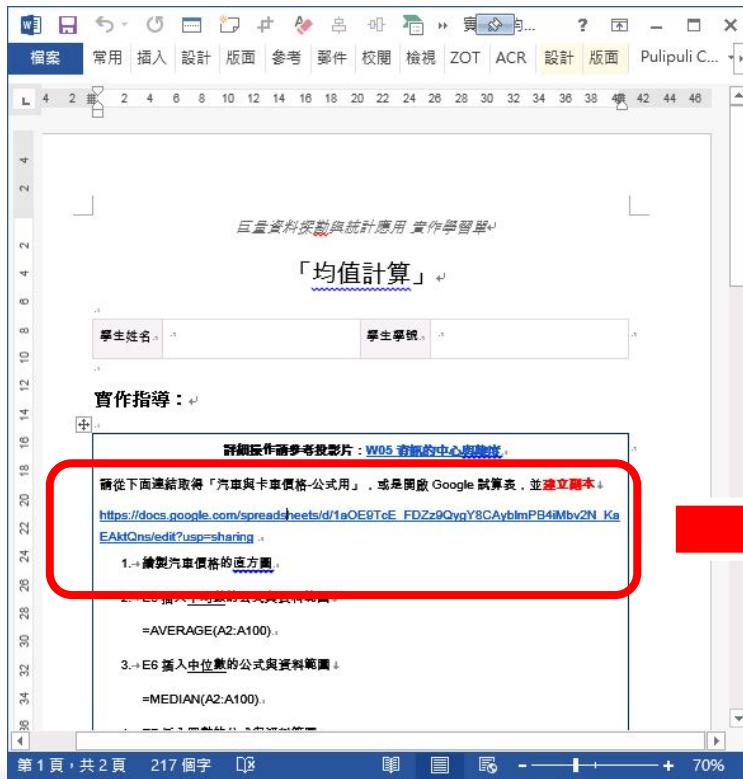
# W05-a. 均值計算

1. 繪製汽車價格的直方圖
2. 插入平均數的公式與資料範圍
3. 插入中位數的公式與資料範圍
4. 插入眾數的公式與資料範圍



# 建立副本

實作:均值計算.docx



巨量資料探索與統計應用 實作學習單

## 「均值計算」

學生姓名	學生學號
------	------

### 實作指導：

詳細實作請參考我影片：[W05 均值的中心與離散](#)

請從下面連結取得「汽車與卡車價格-公式用」，或是開啟 Google 計算表，並建立副本：  
[https://docs.google.com/spreadsheets/d/1aOE9TeE\\_FDZz9QygY8CAyblmPB4iMbv2N\\_KsEAktQns/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1aOE9TeE_FDZz9QygY8CAyblmPB4iMbv2N_KsEAktQns/edit?usp=sharing)。

1.-繪製汽車價格的直方圖。

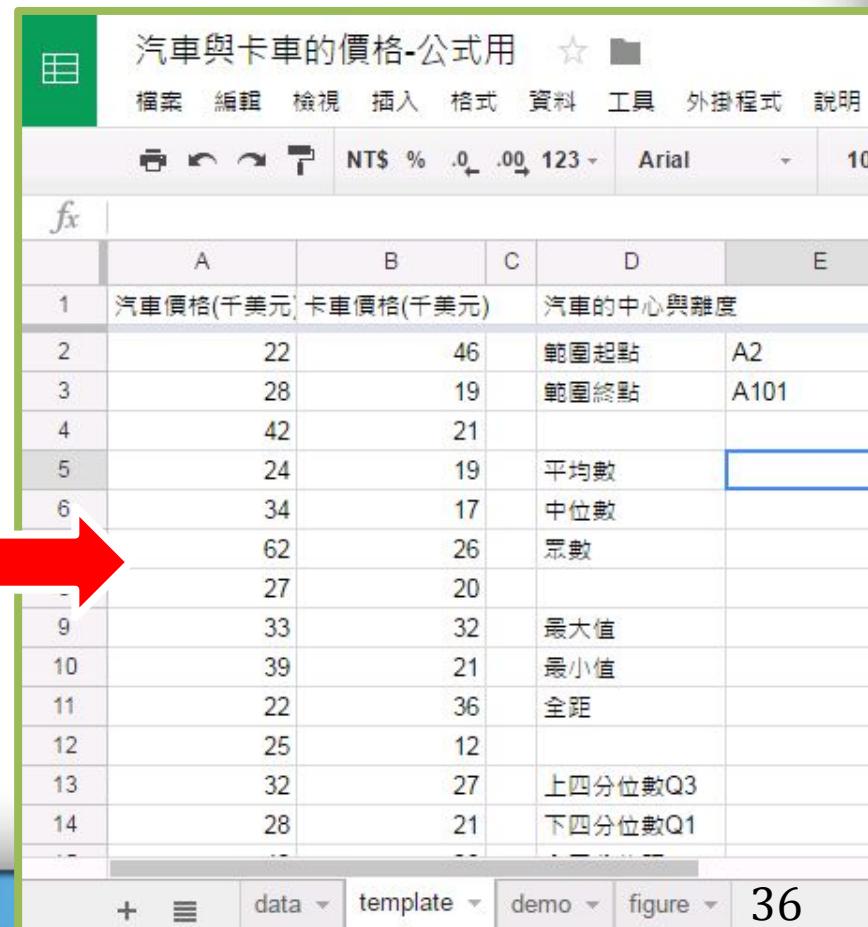
2.-計算汽車與卡車價格的中心與離散。

```
=AVERAGE(A2:A100)
```

3.-E6 插入中位數的公式與資料範圍。

```
=MEDIAN(A2:A100)
```

汽車與卡車的價格-公式用



	A	B	C	D	E
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與離度	
2	22	46		範圍起點	A2
3	28	19		範圍終點	A101
4	42	21			
5	24	19		平均數	
6	34	17		中位數	
	62	26		眾數	
7	27	20			
8	33	32		最大值	
9	39	21		最小值	
10	22	36		全距	
11	25	12			
12	32	27		上四分位數Q3	
13	28	21		下四分位數Q1	
14					

# 1. 繪製汽車價格的直方圖

1. 選擇範圍：  
點選欄名「A」
2. 插入⇒圖表
3. (使用建議圖表)⇒插入
4. 移動圖表到表格空白處



## 2. 插入平均數的公式與資料範圍

1. 點選「平均數」旁邊的空格(E5)
2. 點選「公式列」的空白處
3. 輸入公式：

=AVERAGE(A2:A101)

這是  
公式

公式

資料範圍

2

1

	A	B	C	D	E	F
1	汽車價格(千美元)	卡車價格(千美元)	汽車的中心與離度			
2	22	46	範圍起點	A2		
3	28	19	範圍終點	A101		
4	42	21				
5	24	19	平均數			
6	34	17	中位數			
	62	26	眾數			
7	27	20				
8	33	32	最大值			
9	39	21	最小值			
10	22	36	全距			
11	25	12				
12	32	27	上四分位數Q3			
13	28	21	下四分位數Q1			
14	30	24				

+

data

template

demo

figure

# 資料範圍:單格

A:欄名

2:列名

資料範圍: A2

值: 21.5

2:列名

	A	B
1	汽車價格(千美元)	卡車價格(千美元)
2	22	46
3	28	19
4	42	21
5	24	19
6	34	17
7	62	26

# 資料範圍：範圍多格

起點A2

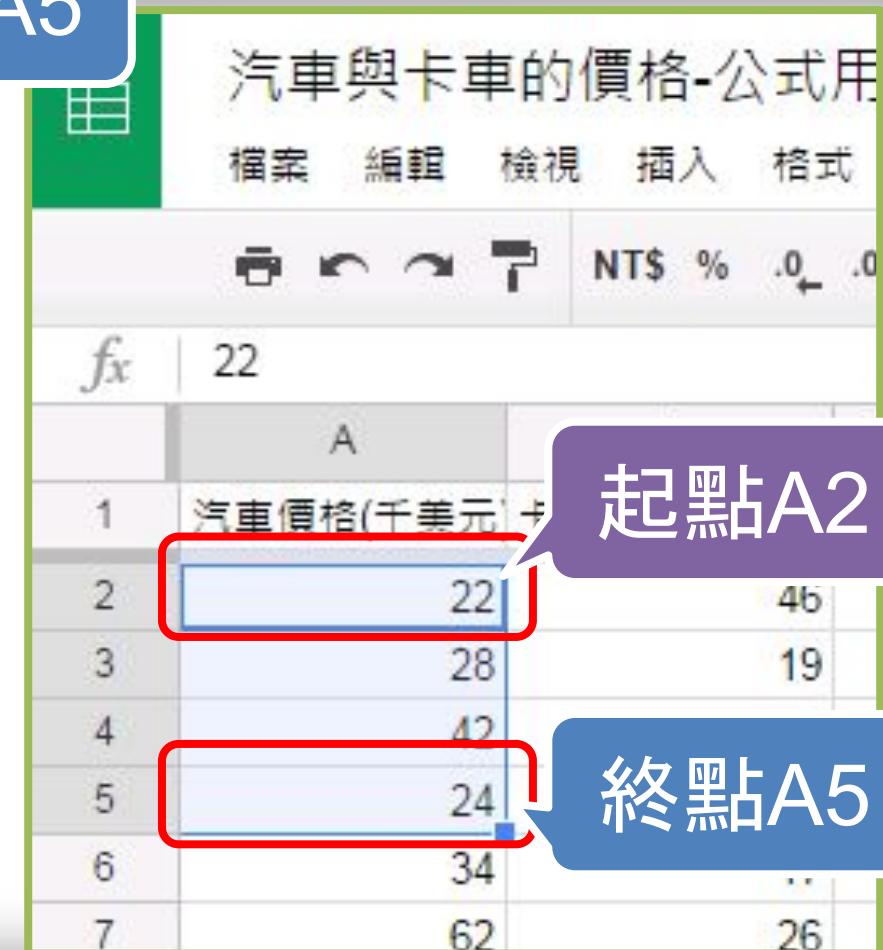
終點A5

資料範圍：**A2:A5**

值：

- 21.5
- 28.4
- 42
- 23.99

範圍分隔：



車款	汽車價格(千美元)	卡車價格(千美元)
1	汽車價格(千美元)	
2	22	45
3	28	19
4	42	
5	24	
6	34	
7	62	26

起點A2

終點A5

### 3. 插入中位數的公式與資料範圍

1. 點選「中位數」旁邊的空格 (E6)
2. 點選「公式列」的空白處
3. 輸入公式：

=MEDIAN(A2:A101)

2

1

	A	B	C	D	E
1	汽車價格(千美元)	卡車價格(千美元)	汽車的中心與離度		
2	22	46	範圍起點	A2	
3	28	19	範圍終點	A101	
4	42	21			
5	24	19	平均數		
6	34	17	中位數		
	62	26	眾數		
7	27	20			
8	33	32	最大值		
9	39	21	最小值		
10	22	36	全距		
11	25	12			
12	32	27	上四分位數Q3		
13	28	21	下四分位數Q1		
14	30	22			

+ data template demo figure

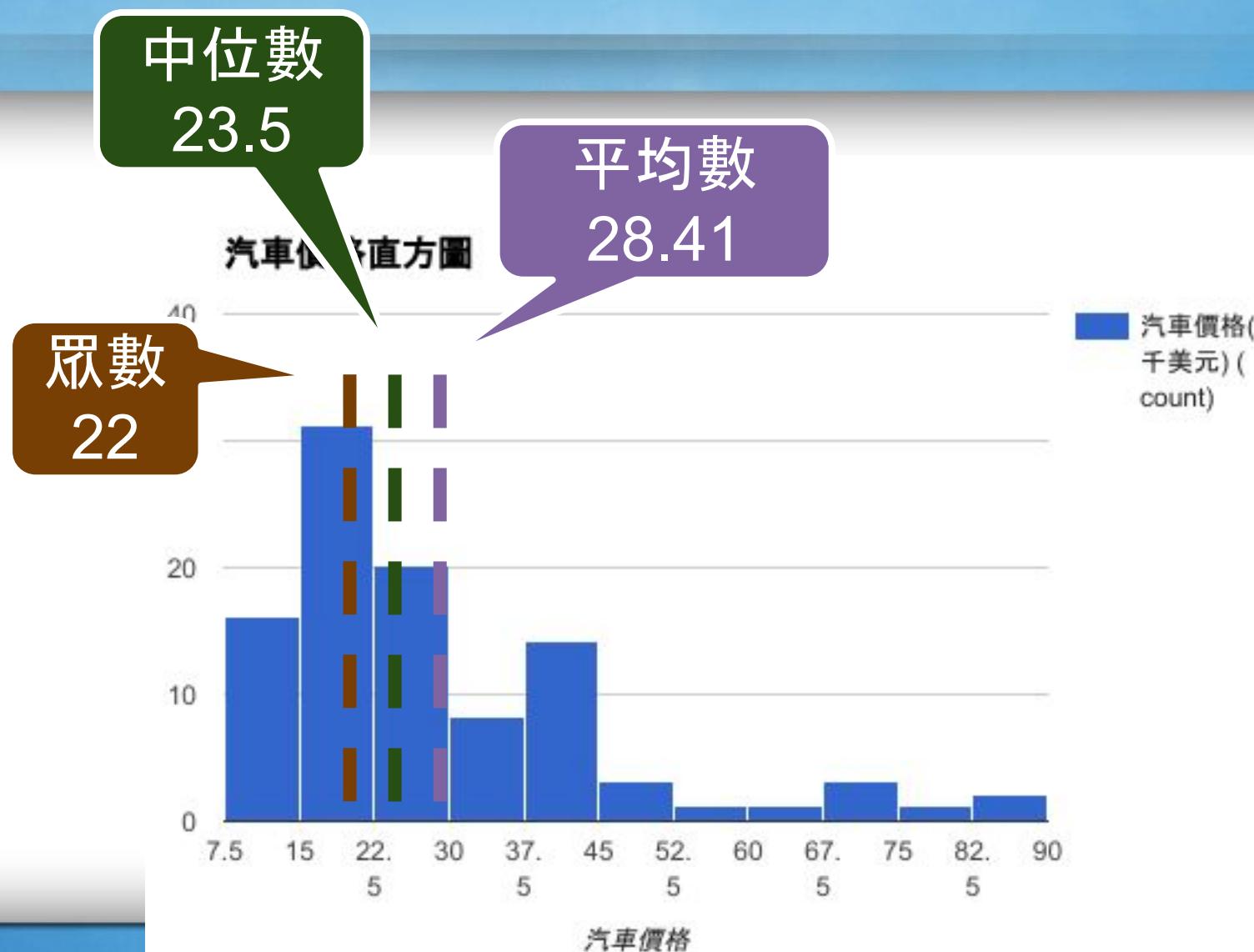
# 4. 插入眾數的公式與資料範圍

1. 點選「眾數」旁邊的空格 (E7)
2. 點選「公式列」的空白處
3. 輸入公式：

=MODE(A2:A101)

The screenshot shows a software interface with a green header bar. The main area is a data table with columns A, B, and C. Row 1 contains the header: '汽車價格(千美元)' in A1, '卡車價格(千美元)' in B1, and '汽車的中心與離度' in C1. Rows 2 through 14 contain data for cars and trucks. Row 15 is a summary row with formulas. The formula bar at the top shows the formula =MODE(A2:A101). A red box highlights the formula bar, and a red speech bubble with the number 2 is positioned above it. A red box highlights the cell E7, and a red speech bubble with the number 1 is positioned to its right. The summary row includes: '範圍起點' (Range Start) in A2, '範圍終點' (Range End) in A101, '平均數' (Mean) in C5 (with a blue box around it), '中位數' (Median) in C6 (with a red box around it), '最大值' (Max) in C7, '最小值' (Min) in C8, '全距' (Range) in C9, '上四分位數Q3' (Upper Quartile Q3) in C10, and '下四分位數Q1' (Lower Quartile Q1) in C11.

# 觀察均值在資料分佈的位置



# W05-a. 均值計算

# 上機實作



## 公式

## 均值

公式 均值

	A	B	C	D
1	汽車價格(千美元)	卡車價格(千美元)	汽車的中	
2	22	46	範圍起點	
3	28	19	範圍終點	A101
4	42	21		
5	24	19	平均數	28.41
6	34	17	中位數	23.5
7	62	26	眾數	22
8	27	20		

Part 3.

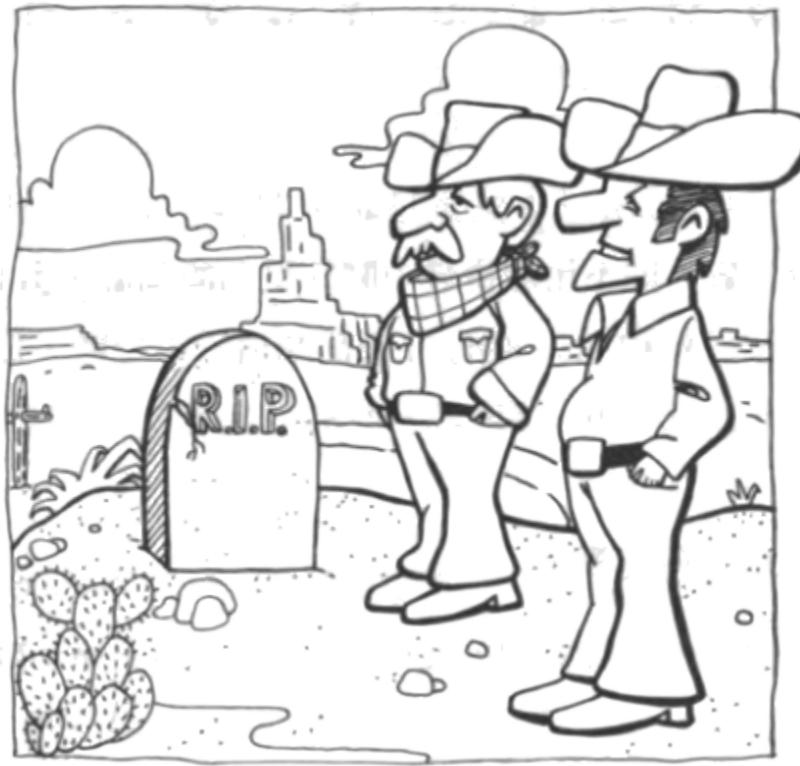
# 資料的離度:分散性

# 離度？那是什麼？



我送你離開 千里之外

# 老王的死因：不懂離度



「他只知道河的**平均深度**，  
居然就打算過河了。」

# 聽說股票很賺



「平均獲利高達**14.62%**，不買是傻子」

# 一年後



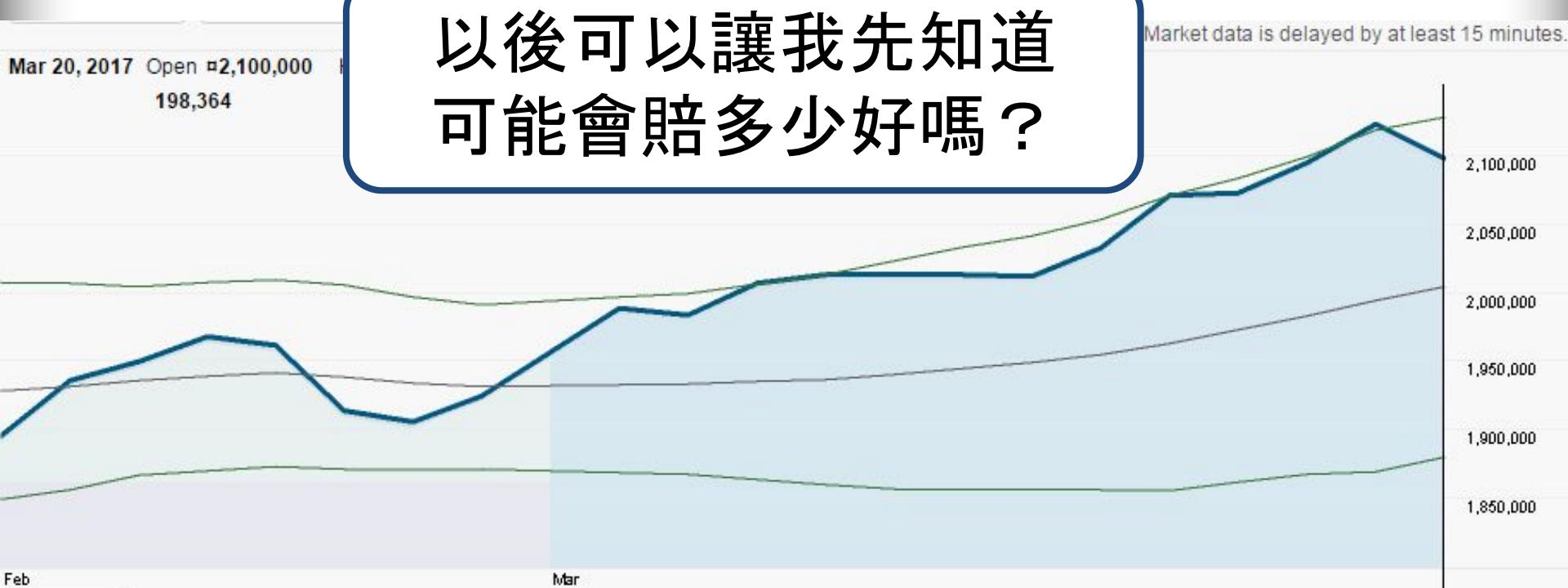
「我虧損了**26%**，怎麼會這樣？」

投資一定有風險，基金投資**有賺有賠**，  
申購前應詳閱公開說明書。

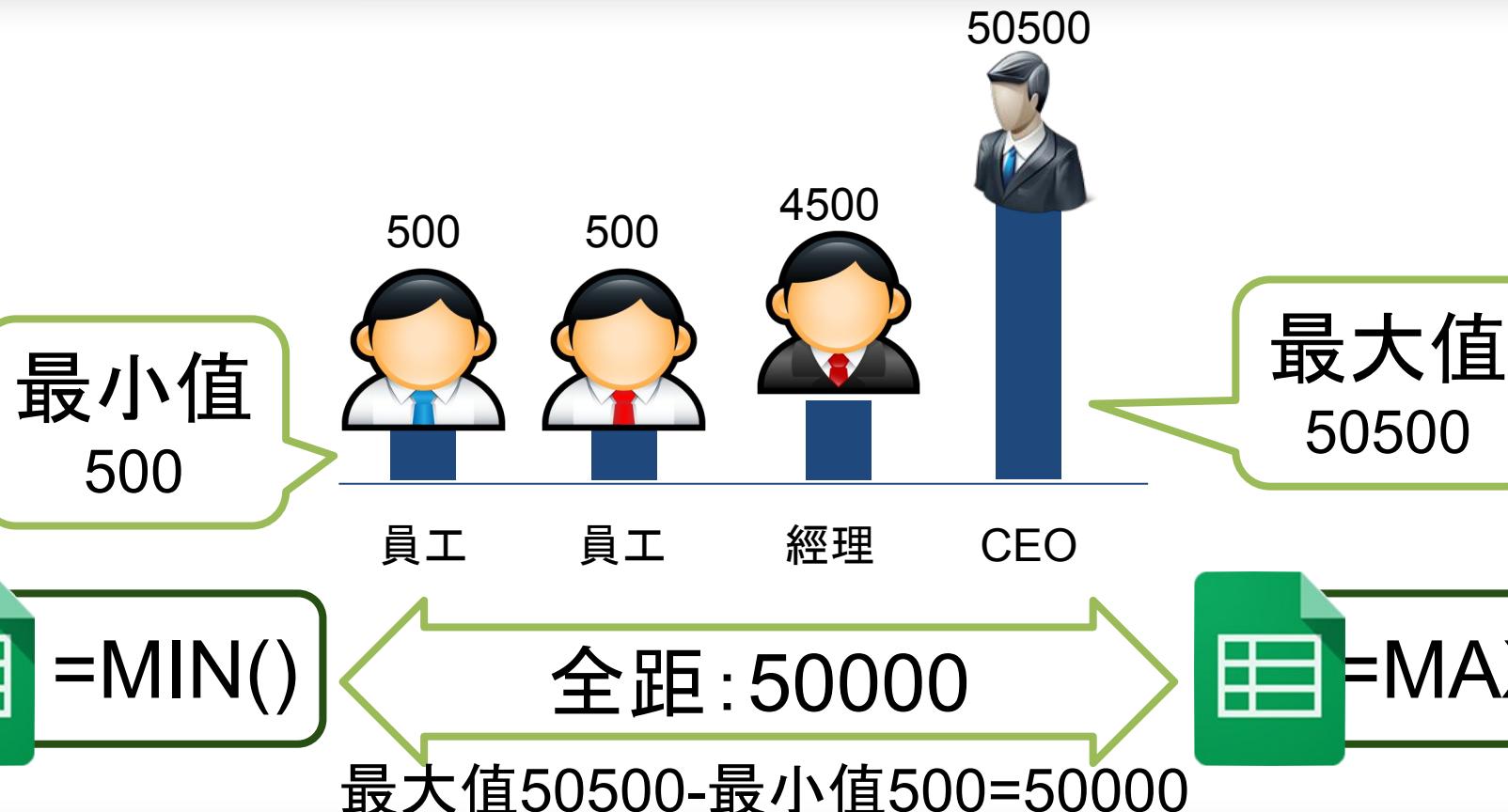




以後可以讓我先知道  
可能會賠多少好嗎？



# 全距:最大值-最小值



# 全距的問題

離群值

全距數：

4000

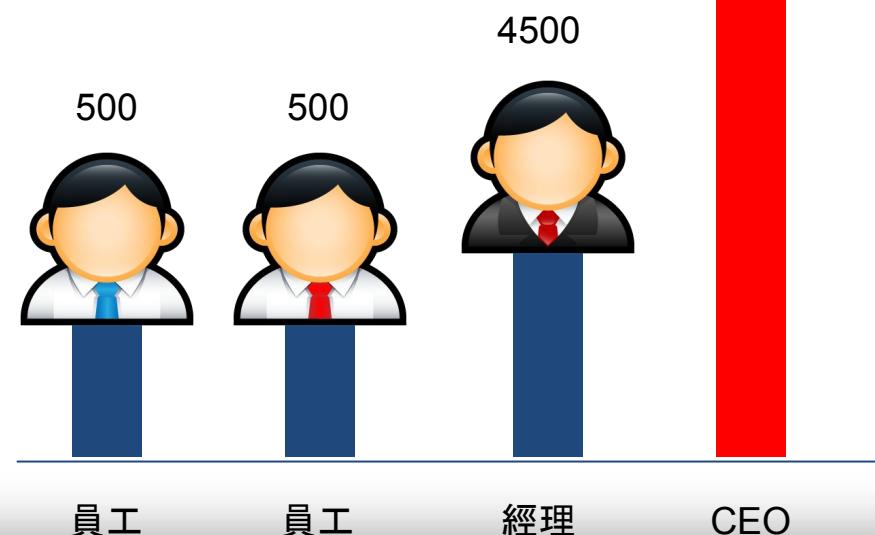
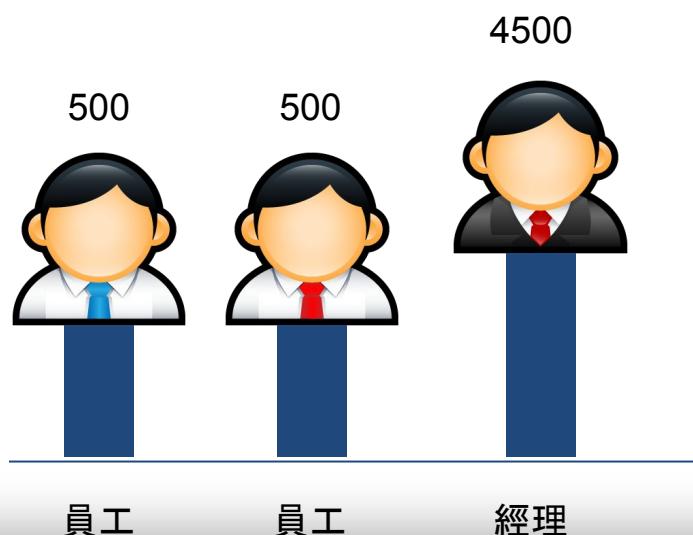


平均數：

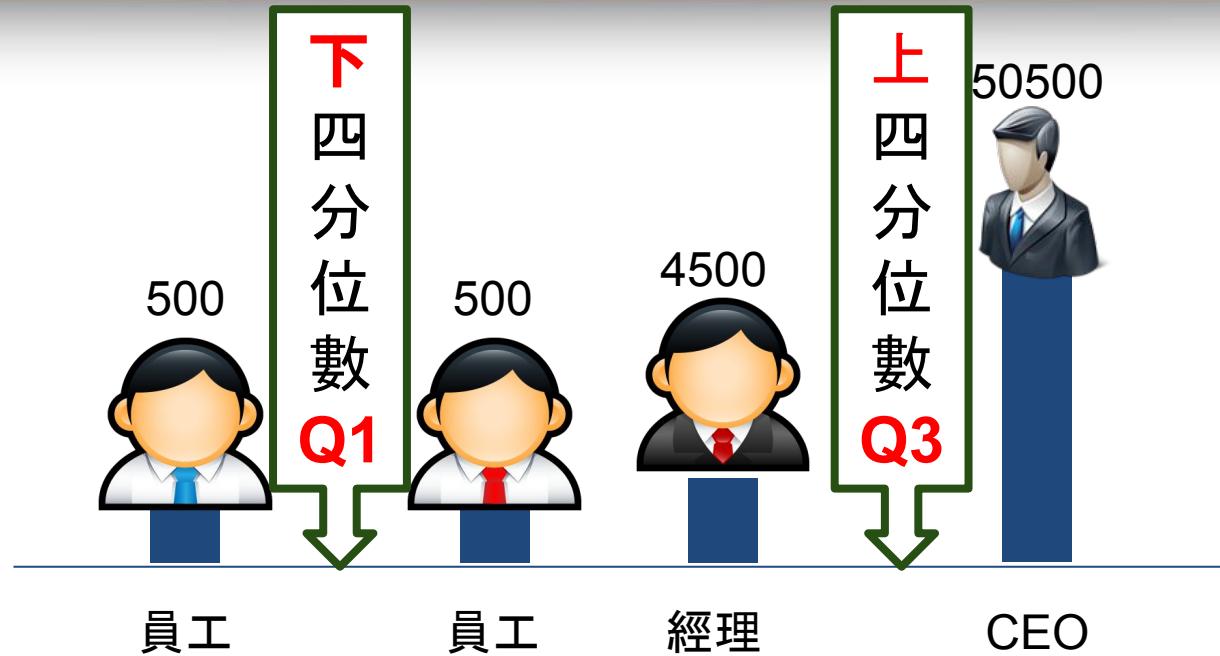
50000

又怪我囉？

50500

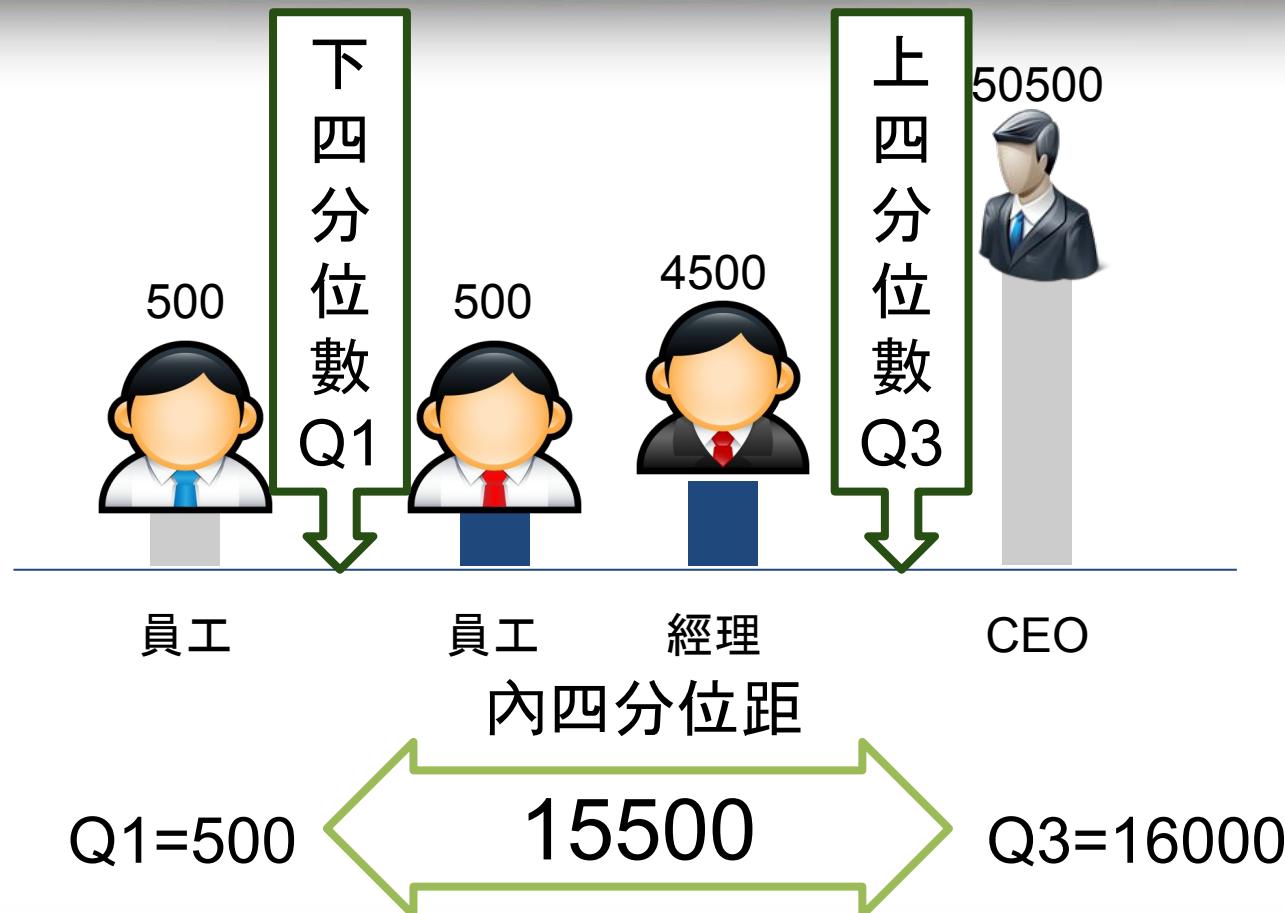


# 內四分位距: 上四分位數-下四分位數



=QUARTILE(資料範圍,1)      =QUARTILE(資料範圍,3)

# 內四分位距: 上四分位數-下四分位數



# 中國動車對撞測試

測試時速:38公里



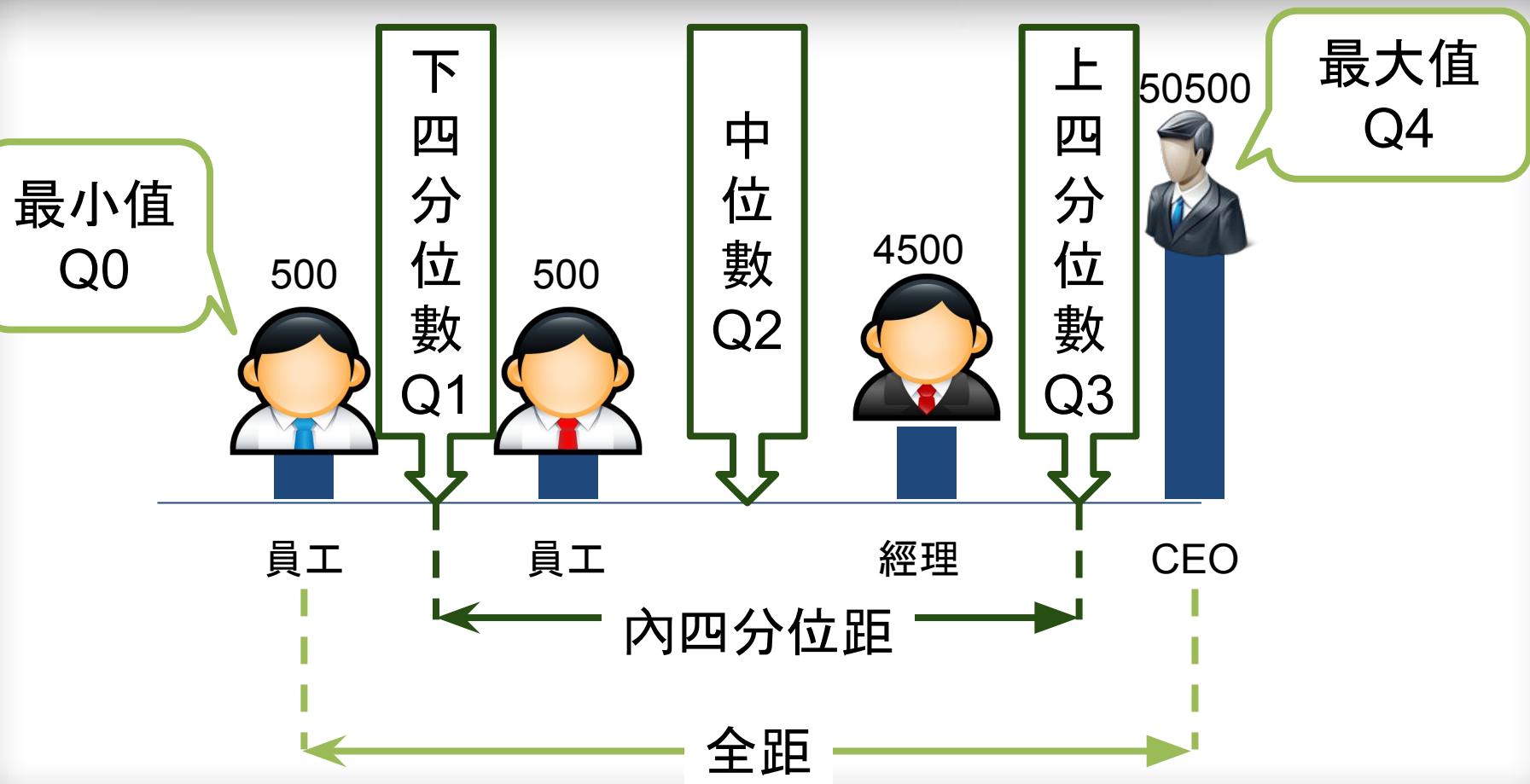
溫州動車對撞40死  
相對時速83公里  
只是特例, 沒事的



離群值

# 五數綜合

Q0, Q1, Q2, Q3, Q4



Part 4.

# 實作:五數綜合的計算

武術 五數綜合聽起來挺強的



「那汽車價格的五數綜合是什麼？」

汽車與卡車...

pulipuli.chen@gmail.com

檔案 編輯 檢視 插入 格式 註解 共用 說明

所有變更都已儲存到雲端硬碟

fx 汽車價格(千美元)

	A	B	C	D
1	汽車價格(千美元)	卡車價格(千美元)		
2	22	46		
3	28	19		
4	42	21		
5	24	19		
6	34	17		
7	62	26		
8	27	20		
9	33	32		
10	39	2		
11	22			
12	25			
13	32			
14	28			
15	40			
16	44			
17	40			

+ data



# 汽車的價格

汽車案例數量: 100筆

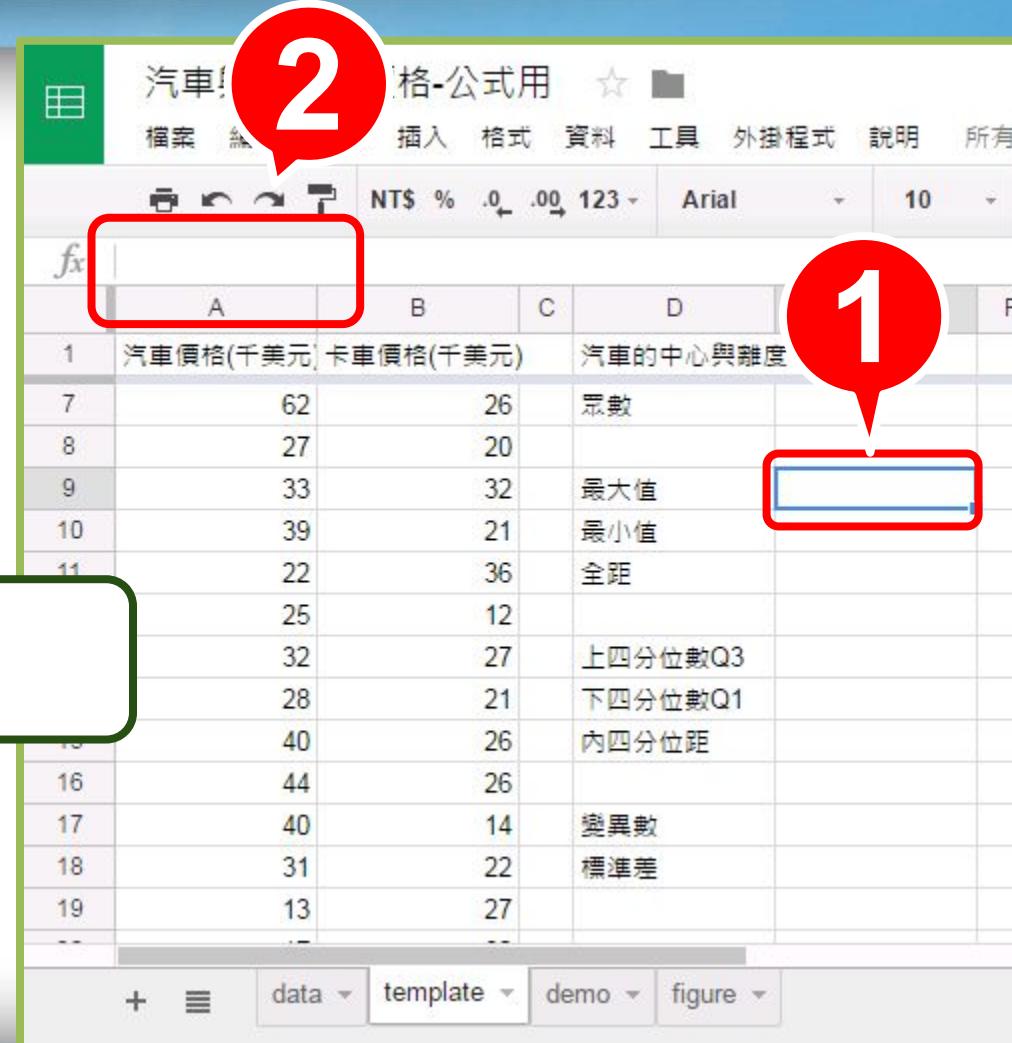
# W05-b. 五數綜合的計算

- 
1. 插入最大值的公式與資料範圍
  2. 插入最小值的公式與資料範圍
  3. 計算全距
  4. 插入上四分位數Q3的公式與資料範圍
  5. 插入下四分位數Q1的公式與資料範圍
  6. 計算內四分位距

# 1. 插入最大值的公式與資料範圍

1. 點選「最大值」旁邊的空格 (E9)
  2. 點選「公式列」的空白處
  3. 輸入公式：

=MAX(A2:A101)



## 2. 插入最小值的公式與資料範圍

1. 點選「最小值」旁邊的空格 (E10)
2. 點選「公式列」的空白處
3. 輸入公式：

$=\text{MIN}(A2:A101)$

The screenshot shows a software interface with a green header bar. The header includes the text '汽車' (Car), '價格-公式用' (Price-Formula Use), and a star icon. Below the header are standard menu options: 插入 (Insert), 格式 (Format), 資料 (Data), 工具 (Tools), 外掛程式 (Add-in), 說明 (Help), and 所有 (All). The main area is a table with two columns of data. The first column is labeled 'A' and the second is labeled 'B'. Row 1 contains the text '汽車價格(千美元)' and '卡車價格(千美元)'. Row 7 contains the value '62'. Row 8 contains the value '27'. Row 9 contains the value '33'. Row 10 contains the value '39'. Row 11 contains the value '22'. Row 25 contains the value '12'. Row 32 contains the value '27'. Row 28 contains the value '21'. Row 40 contains the value '26'. Row 16 contains the value '44'. Row 17 contains the value '40'. Row 18 contains the value '31'. Row 19 contains the value '13'. To the right of the table, there is explanatory text: '汽車的中心與離層' (Center and range of cars), '眾數' (Mode), '最大值' (Maximum), '最小值' (Minimum), '全距' (Range), '上四分位數Q3' (Upper quartile Q3), '下四分位數Q1' (Lower quartile Q1), and '內四分位距' (Interquartile range). A red box highlights the empty cell in the 'E' column of row 10, which is step 1. A red box highlights the formula bar with the formula '=MIN(A2:A101)' and a red number '2' above it, indicating step 2. A red box highlights the cell in the 'E' column of row 10, which is step 3.

### 3. 計算全距

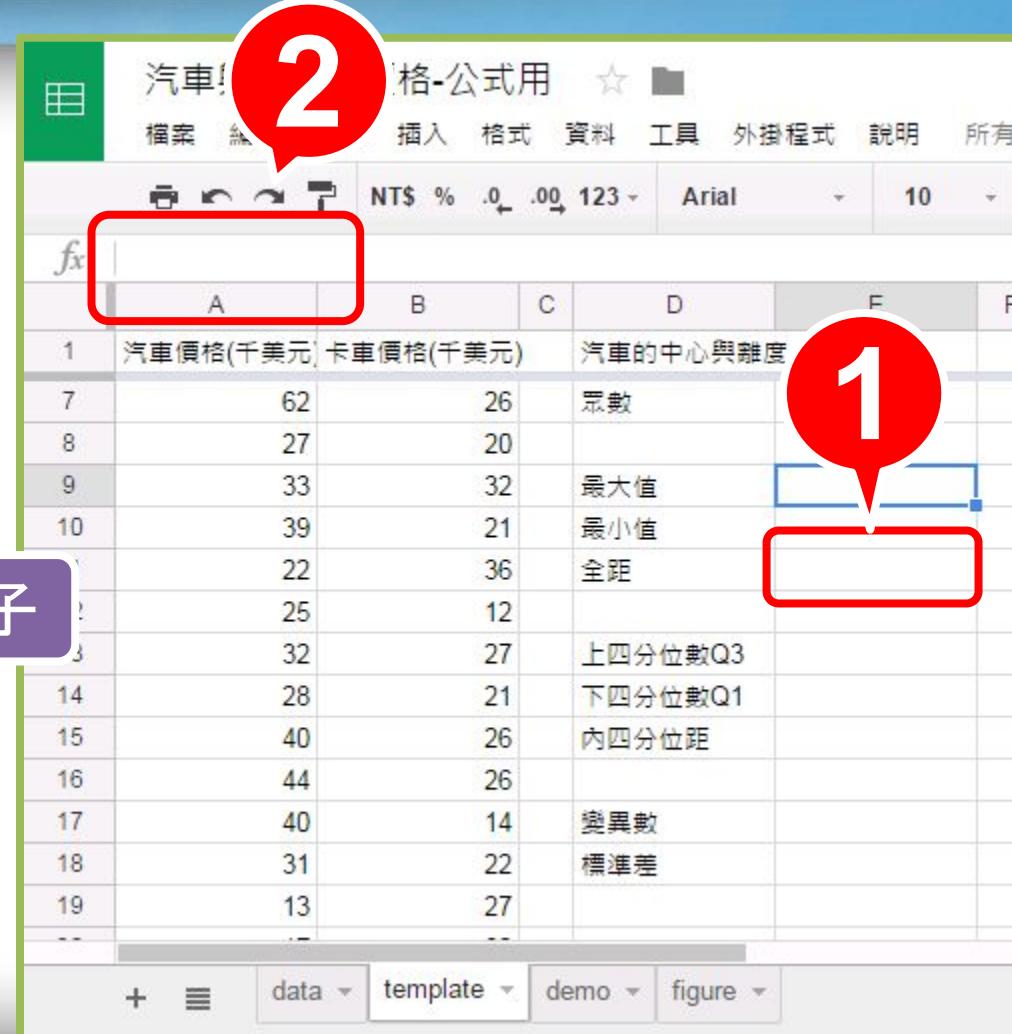
1. 點選「全距」旁邊的空格 (E11)
2. 點選「公式列」的空白處
3. 輸入公式：

最大值的格子

最小值的格子

=E9-E10

運算子(減)



	A	B	C	D	E
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與離度	
7	62	26		眾數	
8	27	20			
9	33	32		最大值	
10	39	21		最小值	
	22	36		全距	
	25	12			
	32	27		上四分位數Q3	
14	28	21		下四分位數Q1	
15	40	26		內四分位距	
16	44	26			
17	40	14		變異數	
18	31	22		標準差	
19	13	27			
--	--	--			

## 4. 插入上四分位數Q3的公式與資料範圍

1. 點選「上四分位數Q3」旁邊的空格 (E13)
2. 點選「公式列」的空白處
3. 輸入公式：

=QUARTILE(A2:A101,3)

The screenshot shows a spreadsheet application with a green header bar. The header bar includes the text '汽車', '價格-公式用', and a star icon. Below the header are standard menu options: 插入, 格式, 資料, 工具, 外掛程式, 說明, and 所有。The main area displays a table with data in columns A, B, and E. Column A contains values 62, 27, 33, 39, 22, 25, 32, 28, 40, 44, 40, 31, 19, and 13. Column B contains values 26, 20, 32, 21, 36, 12, 27, 21, 26, 26, 14, 22, 27, and 27. Column E contains labels: '汽車的中心與離度', '眾數', '最大值', '最小值', '全距', '上四分位數Q3', '下四分位數Q1', '內四分位距', '變異數', and '標準差'. A red box highlights the empty cell E13, which is the target for the formula. A red box also highlights the formula input field 'fx' in the formula bar. A red speech bubble with the number '1' points to the formula input field. A red speech bubble with the number '2' points to the empty cell E13.

# 5. 插入下四分位數Q1 的公式與資料範圍

1. 點選  
「下四分位數Q1」  
旁邊的空格 (E14)
2. 點選「公式列」  
的空白處
3. 輸入公式：

=QUARTILE(A2:A101,1)

The screenshot shows a software interface with a green header bar. The header includes the text '汽車' and '格-公式用' (Formula), along with standard menu options: 插入 (Insert), 格式 (Format), 資料 (Data), 工具 (Tools), 外掛程式 (Add-ins), 說明 (Help), and 所有 (All). The main area is a data table with columns A, B, and E. Row 1 contains the header: '汽車價格(千美元)' in A1, '卡車價格(千美元)' in B1, and '汽車的中心與離度' in E1. Rows 7 through 14 contain data: (62, 26, '眾數'), (27, 20, '最大值'), (33, 32, '最小值'), (22, 36, '全距'), (25, 12, '上四分位數Q3'), (32, 27, '下四分位數Q1'), (28, 21, '內四分位距'), (40, 26, '變異數'), (44, 26, '標準差'). Row 19 contains the formula '=QUARTILE(A2:A101,1)' in the formula bar. A red box highlights the formula bar, and a red callout bubble with the number '2' points to it. A red box highlights the empty cell E14, and a red callout bubble with the number '1' points to it. The bottom of the interface shows a toolbar with buttons for '+', 'data', 'template', 'demo', and 'figure'.

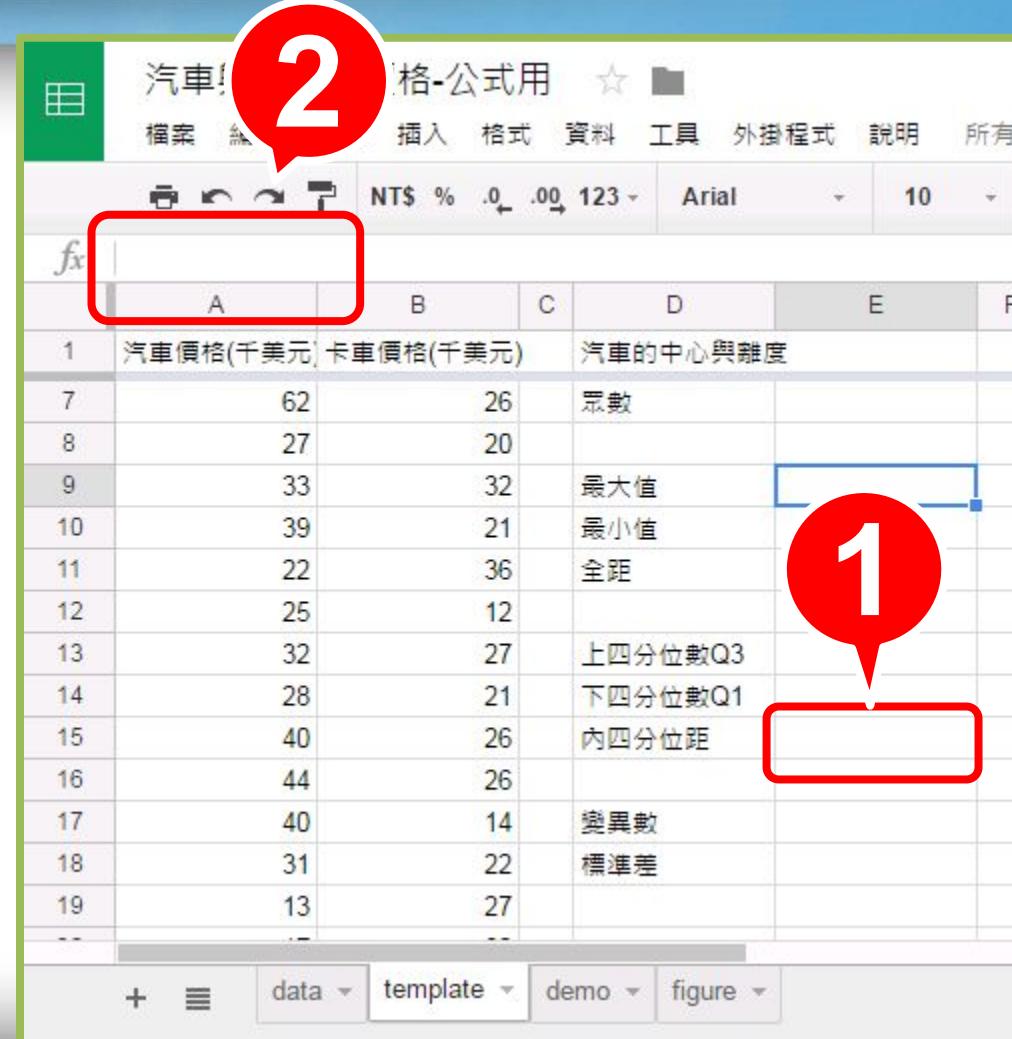
# 6. 計算內四分位距

1. 點選「內四分位距」旁邊的空格 (E15)
2. 點選「公式列」的空白處
3. 輸入公式：

Q3的格子

Q4的格子

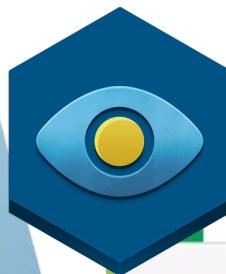
=E13-E14



	A	B	C	D	E
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與離度	
7	62	26		眾數	
8	27	20			
9	33	32		最大值	
10	39	21		最小值	
11	22	36		全距	
12	25	12			
13	32	27		上四分位數Q3	
14	28	21		下四分位數Q1	
15	40	26		內四分位距	
16	44	26			
17	40	14		變異數	
18	31	22		標準差	
19	13	27			
--	--	--			

## W05-b. 五數綜合的計算

# 上機實作



## 公式

# 五數綜合

汽車與卡車的價格...		
1	汽車價格(千美元)	卡車價格(千美元)
7	62	26
8	27	20
9	33	32
10	39	21
		36
		12
15	40	21
16	44	26

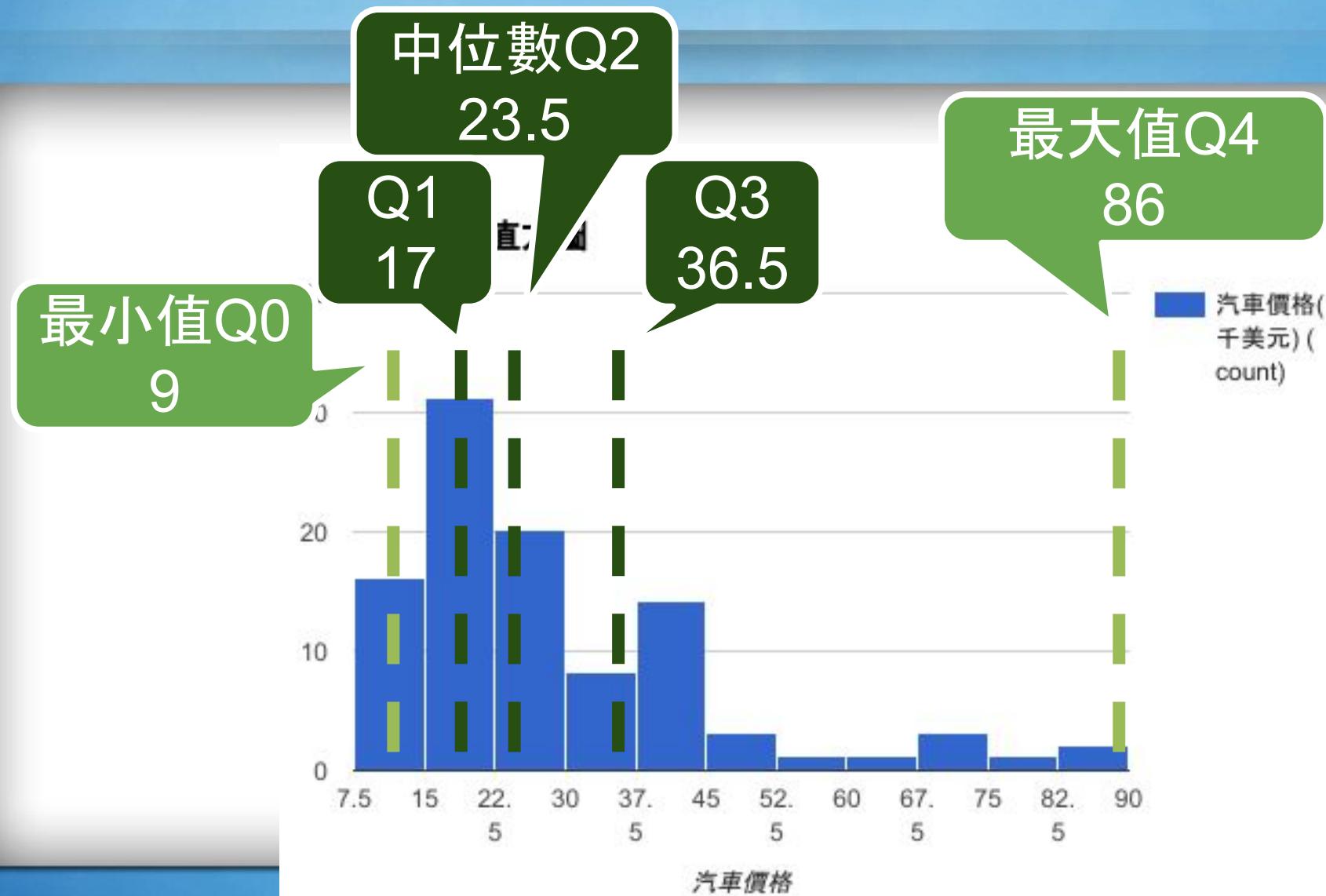
公式

五數綜合

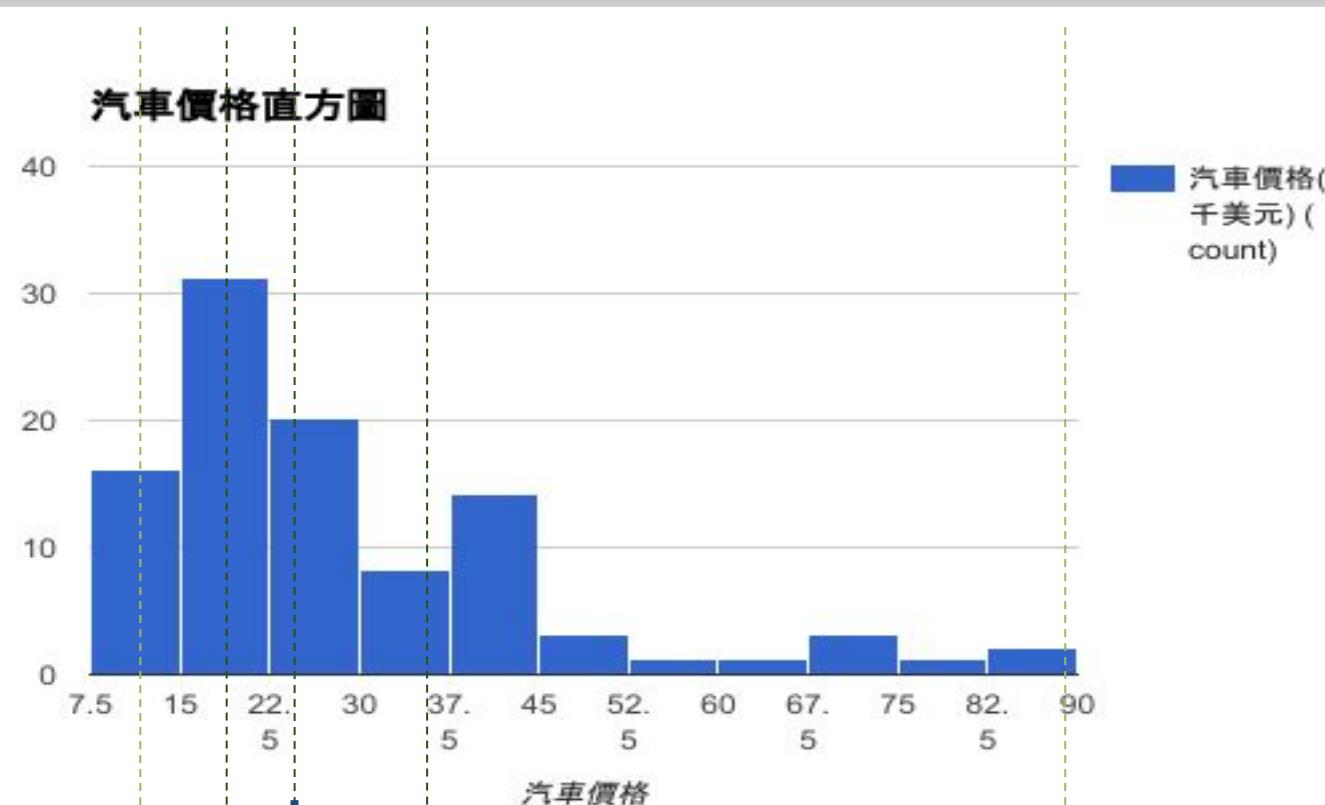
Part 5.

# 實作:五數綜合的繪圖

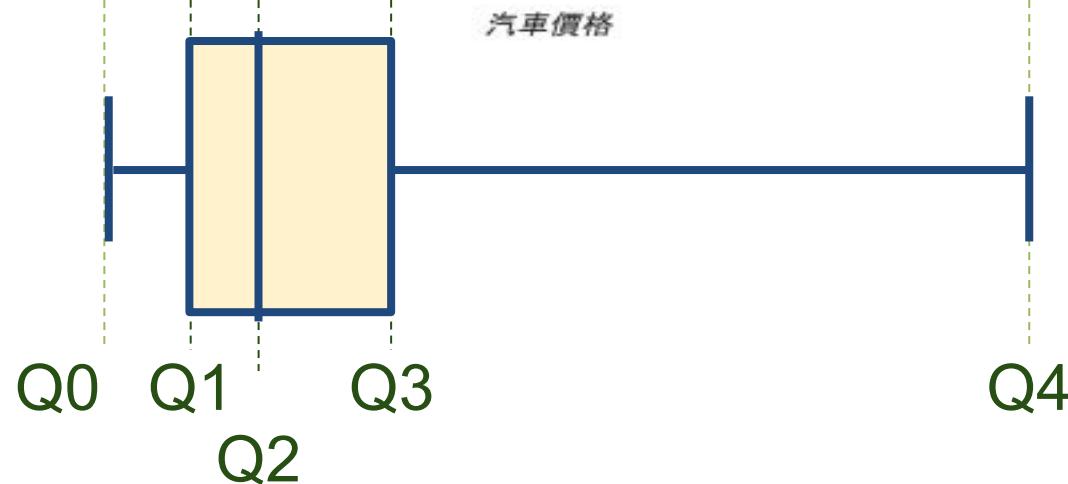
# 觀察五數綜合在資料分佈的位置



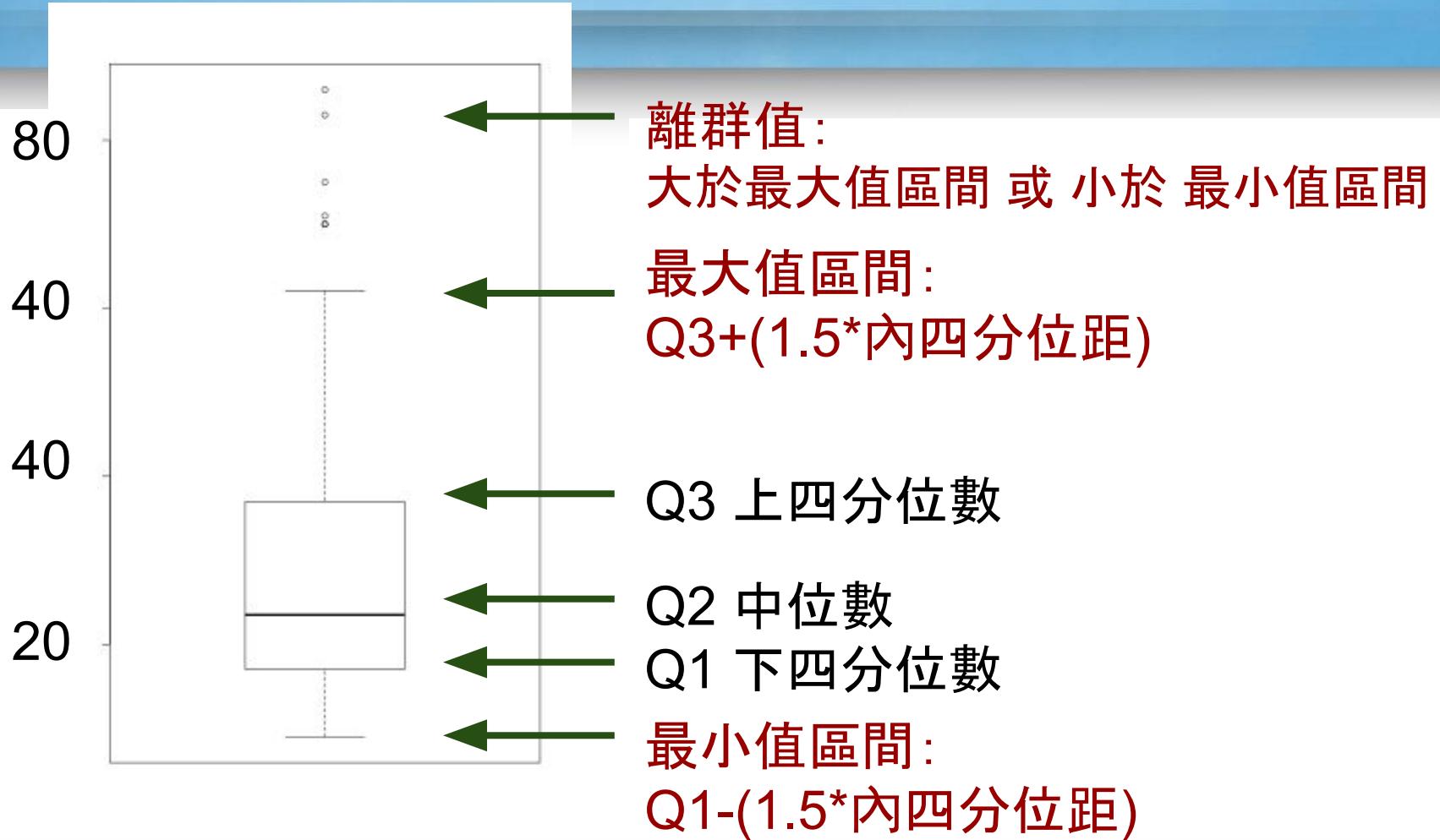
## 直方圖



## 箱型圖 (盒鬚圖)

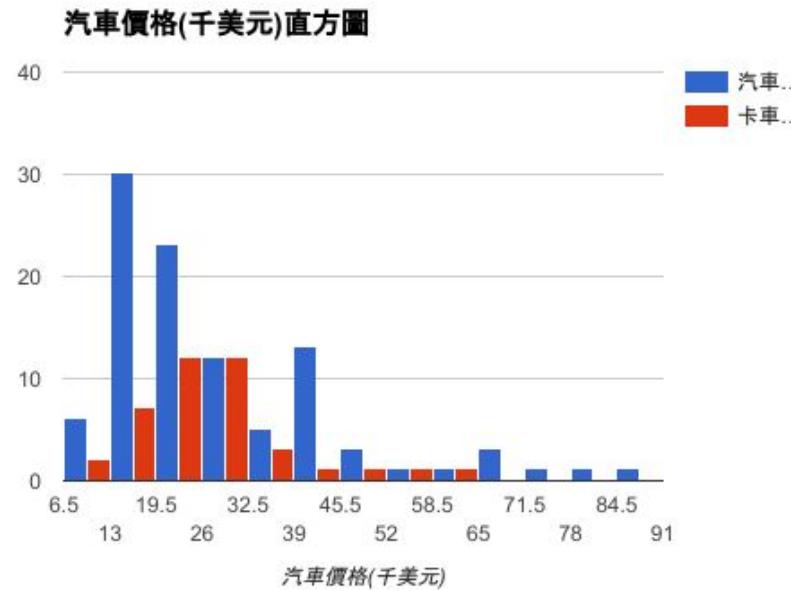


# R的箱型圖(盒鬚圖)

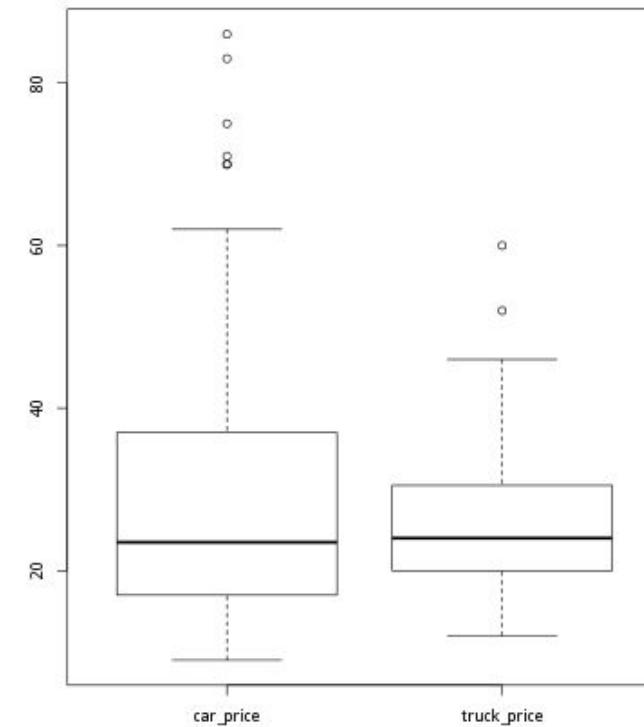


# 箱型圖的特色 雙變項分佈差異比較

## 直方圖



## 箱型圖





# R-Web

## 雲端資料分析暨引導系統

### 統計用程式語言 R

- 主要用於統計分析及統計圖表繪製的程式語言
  - 特別擅長資料處理
  - 多樣擴充套件
- 屬於開放原始碼GNU的項目，跨平臺免費使用
- 沒有圖形化介面，寫錯沒提示，學習曲線陡峭

### 網頁圖形化介面 R-Web

- 以統計入門者為對象，有大量教學
  - 分析技術說明
  - 介面操作錄影
  - 範例檔案可供練習
- 不需安裝，開網頁直接使用
- 處理100筆以上大量資料需要收費

# 費用比較

## R-Web 一般

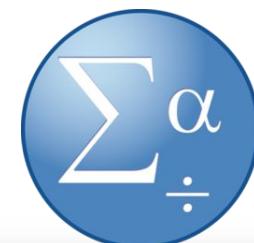
- 免費
- 限制
  - 100筆
  - 5個變數
  - 儲存空間 5MB

## R-Web Base

- 3個月: **\$1,225**
- 免安裝,不限作業系統
- 無資料處理限制
- 可用大部分統計分析方法

## SPSS基本版

- 單機版:**\$96,447**
- 需安裝
- 無資料處理限制
- 可用大部分統計分析方法



有圖當然看圖



「那汽車跟卡車價格的箱型圖是什麼？」

car\_truck\_price

	car_price	truck_price
2	22	46
3	28	19
4	42	21
5	24	19
6	34	17
7	62	26
8	27	20
9	33	32
10	39	21
11	22	36
12	25	12
13	32	27
14	28	21
15	40	26

# 汽車跟卡車的價格

汽車案例數量: 100筆

卡車案例筆數: 40筆



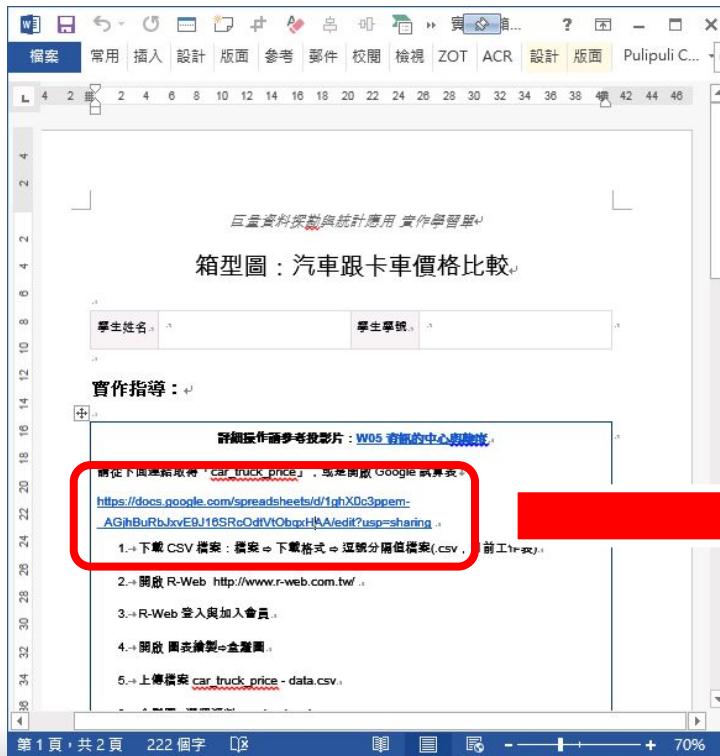
# W05-c. 箱型圖： 汽車跟卡車價格比較



1. 下載CSV檔案
2. 開啟R-Web
3. R-Web登入與加入會員
4. 開啟 圖表繪製⇒盒鬚圖
5. 上傳檔案
6. 盒鬚圖⇒選擇資料
7. 選擇變數
8. 取得盒鬚圖(箱型圖)

# 開啟檔案

實作：箱型圖：汽車跟卡車價  
格比較.docx



實作：箱型圖：汽車跟卡車價格比較

學生姓名	學生學號
------	------

實作指導：

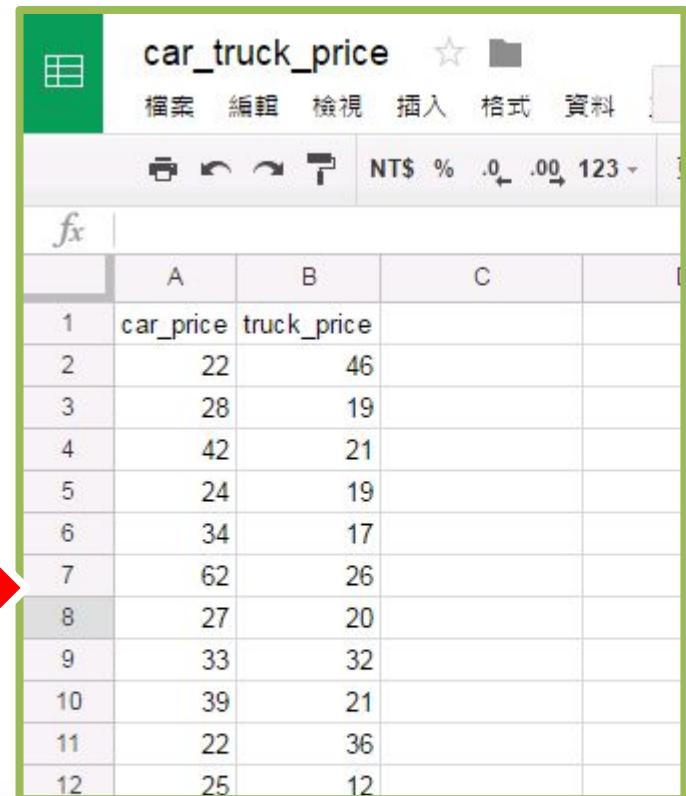
詳解實作請參考我影片：[WPS 實驗的中心與點](#)

請從下面連結取得「car\_truck\_price」，或是開啟 Google 請算表。

<https://docs.google.com/spreadsheets/d/1ghX0c3ppem-AGihBuRbJxvEBJ185RcOdfVtObpxt4LA/edit?usp=sharing>

- 1.→ 下載 CSV 檔案：檔案 → 下載格式 → 逗號分隔值檔案(.csv)
- 2.→ 開啟 R-Web <http://www.r-web.com.tw/>
- 3.→ R-Web 登入與加入會員。
- 4.→ 開啟 圖表繪製-盒鬚圖。
- 5.→ 上傳檔案 car\_truck\_price - data.csv。

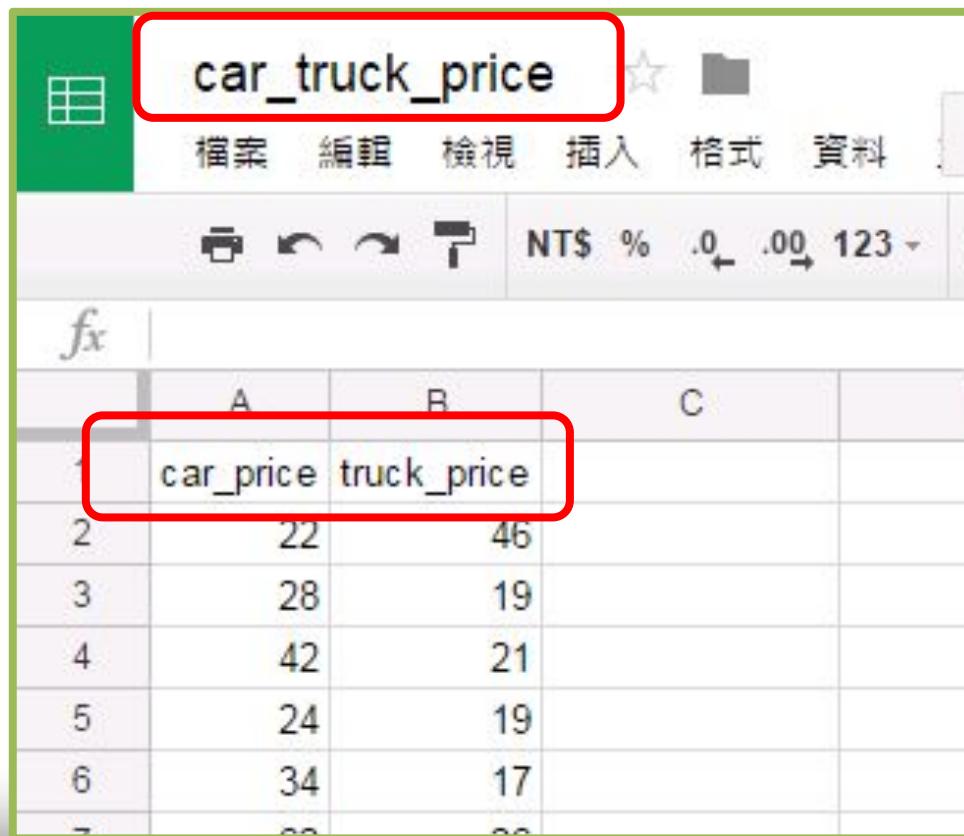
car\_truck\_price



	A	B
1	car_price	truck_price
2	22	46
3	28	19
4	42	21
5	24	19
6	34	17
7	62	26
8	27	20
9	33	32
10	39	21
11	22	36
12	25	12

# R-Web命名限制

英文、數字、\_ (底線) 、 - (橫線)

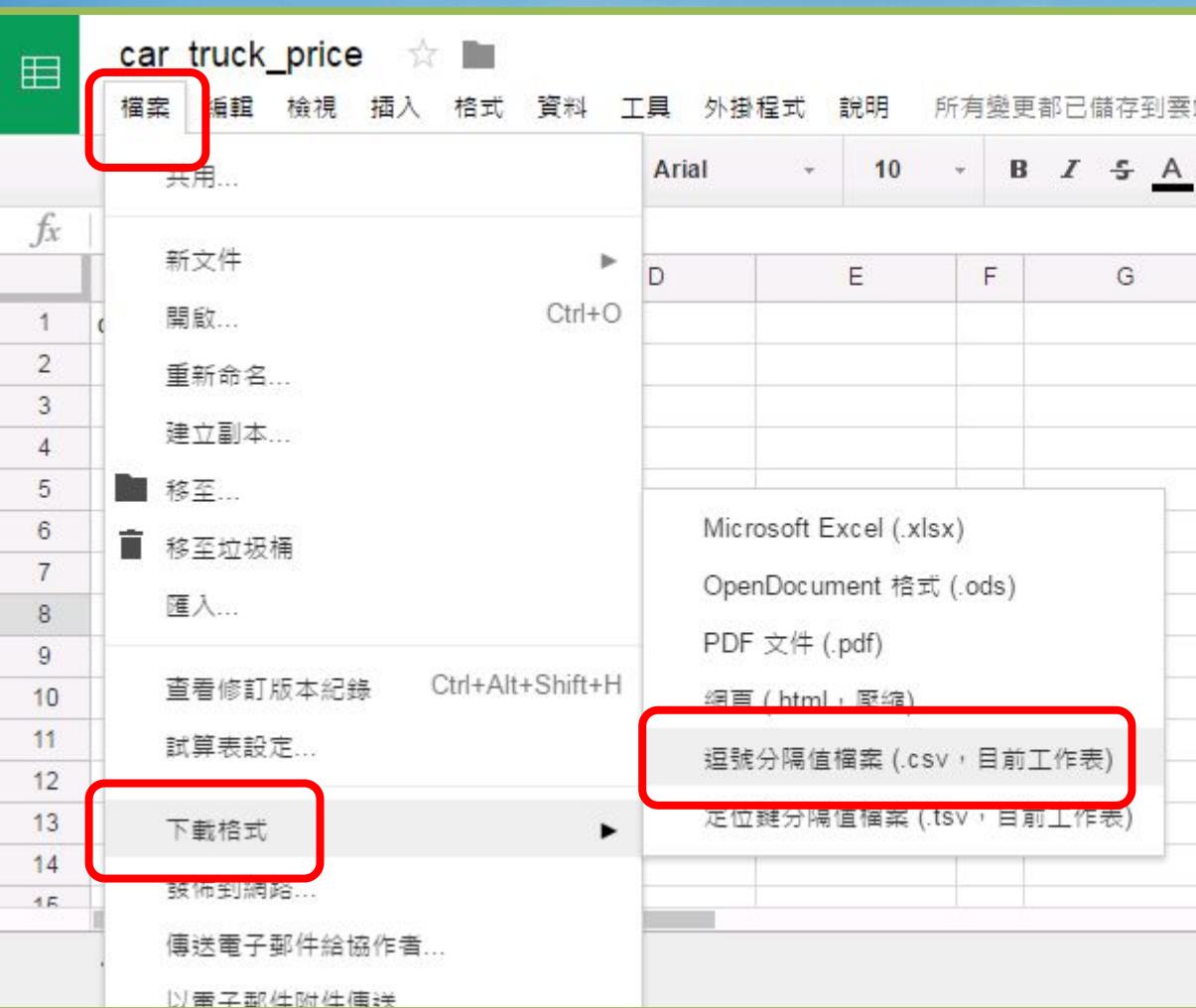


The screenshot shows a Microsoft Excel spreadsheet with the following details:

- Title Bar:** The title "car\_truck\_price" is displayed in the top-left corner, with the first part "car\_truck\_" highlighted by a red box.
- Toolbar:** Standard Excel toolbar with icons for print, undo, redo, and various data formats.
- Formulas Bar:** The text "fx" is visible in the formula bar.
- Table:** A data table with columns labeled "A" and "B". The first row contains the column headers "car\_price" and "truck\_price", both of which are highlighted by a red box.
- Data Rows:** Rows 2 through 7 contain the following data:

	A	B
1	car_price	truck_price
2	22	46
3	28	19
4	42	21
5	24	19
6	34	17
7	22	22

# 1. 下載CSV檔案



1. 檔案
2. 下載格式
3. 逗號分隔值檔案 (.csv, 目前工作表)



car\_truck\_price -  
data.csv

# 2. 開啟R-Web

**R-Web**  
<http://www.r-web.com.tw/>

雲端資料分析暨導引系統  
Data Analysis & Guiding System - Cloud

首頁 | 網站導覽 | 會員登入 | 初階使用者  
Facebook 分享 G+1

初階導引 資料處理 分析方法 圖表繪製 機率分配 輸出結果 自創巨集

**使用流程簡介**

此流程圖為提供新手/初階使用者  
使用本系統之建議順序，若為專家  
使用者可直接選取所需功能進行分析

**DAGS-C**

資料未處理  
資料已處理  
資料處理功能  
系統導引功能  
分析方法功能  
建議程序  
進階程序  
輸出結果功能

雲端資料分析暨導引系統(Data Analysis & Guiding System-Cloud, DAGS-C)  
在網路上以視窗點選方式操作，資料處理及分析零負擔。

最新消息  
雲端計算聯盟  
加入會員

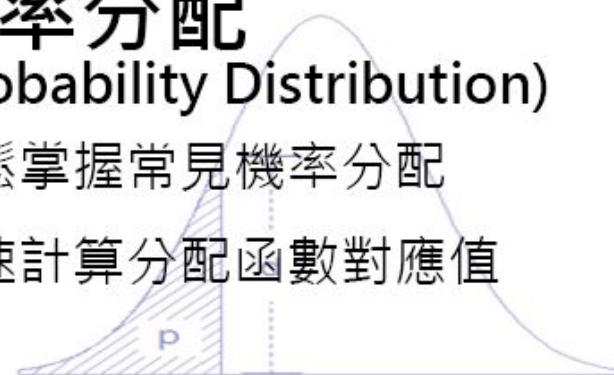
### 3. R-Web登入與加入會員 (1/3)

首頁 ⇨ 會員登入

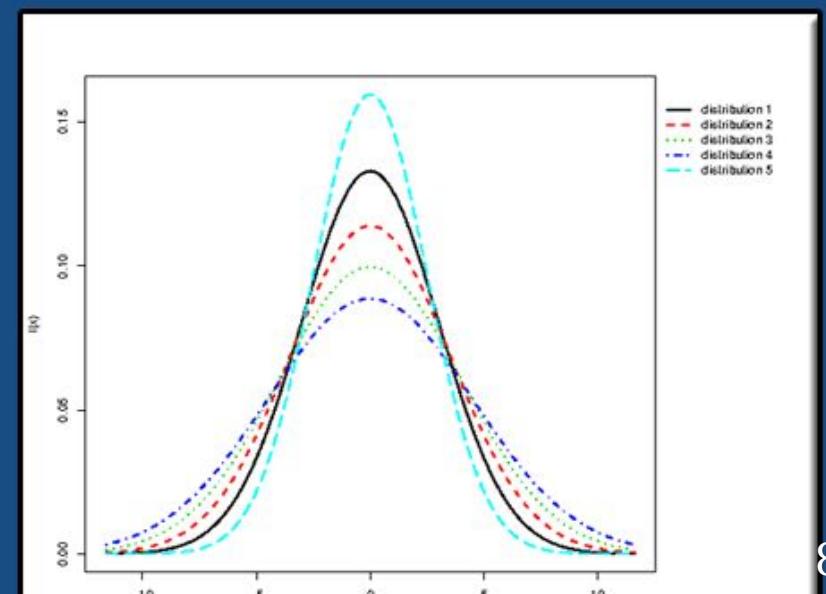


## 機率分配 (Probability Distribution)

輕鬆掌握常見機率分配  
快速計算分配函數對應值



**DAGS-C**



### 3. R-Web登入與加入會員 (2/3)

登入或加入會員

首頁 » 會員登入

會員登入  
Log In

登入

登入

帳號 :

密碼 :

登入

忘記密碼？

加入會員

- 本功能僅限會員使用，請先行登入或加入會員

# 3. R-Web登入與加入會員 (3/3)

## 加入會員 (免費)

1. 確認服務條款
2. 填寫會員資料
  - a. 會員帳號:4~12字
  - b. 會員密碼:8~12字
  - c. 電子信箱
  - d. 西元出生年
  - e. 性別
  - f. 行業
  - g. 工作或居住地區
3. 信箱驗證
4. 完成註冊
5. 登入



The screenshot shows the 'Join RDAGS Member' registration page. At the top, there is a logo of a stylized 'R' and the text '雲端資料分析暨導引系統' (Cloud Data Analysis & Guiding System). Below the logo, there are several navigation links: 首頁, 網站導覽, 會員登入, 初階導引, 資料處理, 分析方法, 圖表繪製, 機率分配, 輸出結果, and 自創巨集. A circular graphic on the right side indicates a process flow: 會員資料 (Membership Information) leads to 服務條款 (Service Terms), which leads to 信箱驗證 (Email Verification).

The registration form is divided into two main sections:

- 一、會員帳號資料**

會員帳號 Member ID	4~12字元，且不含特殊字元
會員密碼 Password	8~12字元，且不含特殊字元
確認密碼 Confirm Password	請再次輸入密碼
- 二、個人基本資料**

電子郵件 E-mail Address	系統將寄發驗證信至此信箱，一組電子郵件帳號僅可申請一個會員帳號
西元出生年 Year of Birth	請選擇
性別	男 ♂ 女 ♀

A note at the bottom of the second section states: '所有欄位皆為必填資料' (All fields are required).

# 4. 開啟 圖表繪製→盒鬚圖

首頁 ⇒ 圖表繪製 ⇒ 盒鬚圖

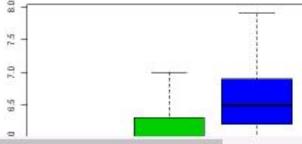
The screenshot shows the 'Cloud-based Data Analysis & Guiding System' (DAGS.C) interface. The top navigation bar includes links for '首頁' (Home), '網站導覽' (Website Navigation), '會員登入' (Member Login), and '初階' (Beginner). The main menu has items: '初階導引' (Beginner Guide), '資料處理' (Data Processing), '分析方法' (Analysis Methods), '圖表繪製' (Chart Drawing), '機率分配' (Probability Distribution), '輸出結果' (Output Results), and '自創巨集' (Create Your Own). The '圖表繪製' (Chart Drawing) item is highlighted with a red box. A dropdown menu on the right lists various chart types: '次數分配表' (Frequency Distribution Table), '列聯表' (Contingency Table), '莖葉圖' (Stem-and-leaf Plot), '散佈圖' (Scatter Plot), '曲線(面)圖' (Curve(Face) Plot), '直方圖' (Histogram), '長條圖' (Bar Chart), '圓餅圖' (Pie Chart), and '盒鬚圖' (Box Plot). The '盒鬚圖' (Box Plot) item is also highlighted with a red box and has a red arrow pointing to it. The central area displays a scatter plot titled 'Scatter plot for 345 rules' with a color scale from red to blue. The bottom navigation bar includes the DAGS.C logo and the URL 'www.r-web.com.tw/chart/step1.php?method=box\_chart'. The page title is '資料採礦(Data Mining) 關聯規則分析-Apriori method'.

# 5. 上傳檔案 (1/5)

盒鬚圖  
Box and Whisker Plot

方法簡介

盒鬚圖，又稱為箱型圖，是一種檢視(比較)資料分散程度時相當實用的一種統計圖。盒鬚圖中會顯示資料的最大值、最小值、中位數、第一四分位以及第三四分位數等統計量。



步驟一：資料匯入

選擇要進行分析的資料檔或 [上傳檔案](#)

== 請選擇一個資料夾 == ▾ 檢視資料型態(開啟視窗)  
== 請選擇一個資料夾 ==

您所選擇的資料檔為： == 尚未選擇資料檔 ==

# 5. 上傳檔案 (2/5)

首頁 » 資料處理 » 管理資料檔 » 上傳資料檔

## 上傳資料檔 Upload Data Files

方法簡介 范例說明 影音教學

無論是欲分析的資料或是輔助分析的額外資料，  
由此功能皆可輕鬆上傳，並佔用系統所提供的個人使用空間。  
本系統目前可提供之檔案格式包含：

步驟一：選擇資料檔

- 請選擇要上傳的資料檔：

car\_truck\_price - data.csv

  
car\_truck\_price -  
data.csv

確認上傳

Rights Reserved

最佳瀏覽解析度 1280\*800 以上 / Chrome

# 5. 上傳檔案 (3/5)

首頁 » 資料處理 » 管理資料檔 » 上傳資料檔

## 上傳資料檔 Upload Data File

步驟二：上傳資料設定

- 變數名稱是否包含在資料檔的最頂端：  
 是  否

資料中的變數是否包含在最頂端：



5.	34	17
6.	62	26

回上一步 下一步驟



下一步驟

# 5. 上傳檔案 (4/5)

上傳資料檔  
Upload Data File

步驟三：修改資料名稱及變數型態

欲將資料檔加密，請輸入密碼：

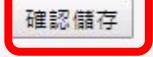
資料檔名稱  

變數名稱

變數型態  

	22	46
1.		
2.	28	19
3.	42	

\* 使用者若對資料有隱私、安全性上之考量，建議可將資料檔名稱以代碼取代。

回上一步  確認儲存

輸入資料檔案名稱  
car\_truck\_price  
英數字,底線/橫線

確認儲存

雲端資料分析暨導引系統 © 2013 All Rights Reserved

最佳瀏覽解析度 1280\*800 以上 / Chrome

# 5. 上傳檔案 (5/5): 完成

雲端資料分析暨導引系統  
Data Analysis & Guiding System - Cloud

首頁 | 網站導覽 | 會員專區/登出 | 初階使用者 ▾

f 誉 分享 G+1

初階導引 資料處理 分析方法 圖表繪製 機率分配 輸出結果 自創巨集

首頁 » 管理資料檔 » 資料檔列表 » 使用者個人資料檔

資料檔列表 Data Files

使用者個人資料檔列表

回去  
盒鬚圖

	資料檔名稱	檢視	編輯	加密	觀測值(列)個數	變數(行)個數	檔案大小	最後修改日期
<input type="checkbox"/>	0320_car_truck				100	2	1.2 K	2017-03-20
<input type="checkbox"/>	14				23	2	675.0 B	2017-02-15
<input type="checkbox"/>	15				19	3	1.2 K	2017-02-15
<input type="checkbox"/>	24_stu_smoke				2	3	590.0 B	2017-02-15
<input type="checkbox"/>	PGC				18	2	562.0 B	2017-02-16

# 6. 盒鬚圖→選擇資料

使用者個人資料檔

選擇要進行分析的資料檔或上傳檔案

剛剛上傳的檔案

使用者個人資料檔 檢視資料型態(開啟視窗)

PCS3  
car\_sales  
car\_truck\_price  
car\_truck\_price\_2  
car\_truck\_price\_int

您所選擇的資料檔為： car\_truck\_price

步驟二：參數設定

選擇欲繪製盒鬚圖之變數

選擇分類變數

(若選用此項則僅能繪製單一變數盒鬚圖)

car\_price  
truck\_price

不選用分類變數 ▼

繪製圖形

進階選項

重新設定

# 7. 選擇變數

## 步驟一：資料匯入

選擇要進行分析的資料檔或[上傳檔案](#)

使用者個人資料檔

檢視資料型態(開新視窗)

IPCS3

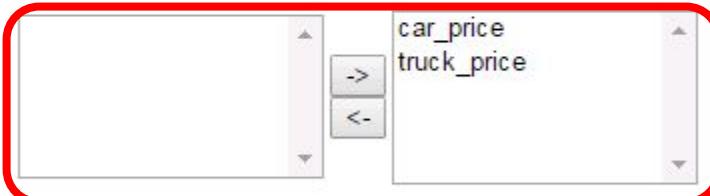
選取car\_price跟  
truck\_price  
-> 到右邊

## 步驟二：參數設定

選擇欲繪製盒鬚圖之變數

選擇分類變數

(若選用此項則僅能繪製單一變數盒鬚圖)



不選用分類變數 ▼

繪製圖形

繪製圖形

進階選項

重新設定

# 8. 取得盒鬚圖(箱型圖)

自頁 » 輸出結果 » 統計圖表：盒鬚圖

## 輸出結果

Output

盒鬚圖 - car\_truck\_price

- 資料名稱：car\_truck\_price
- 變數名稱：car\_price, truck\_price
- 計算時間：0.029秒

- 變數名稱：car\_price, truck\_price
- 計算時間：0.029秒

• 盒鬚圖：

car\_price

truck\_price

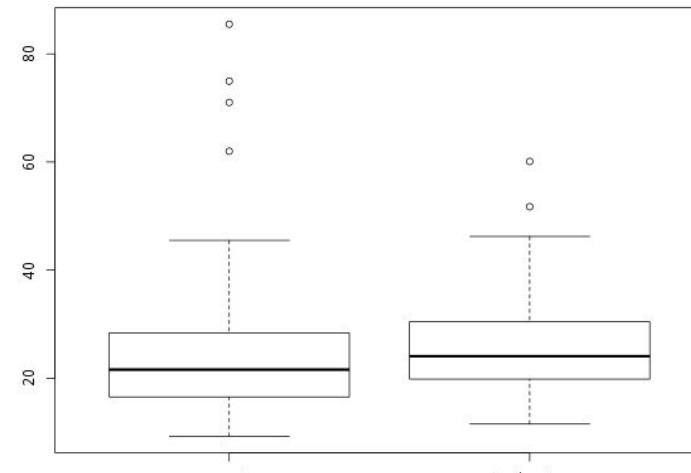
# W05-c. 箱型圖： 汽車跟卡車價格比較

上機實作

參數

箱型圖

- 資料名稱 : car\_truck\_price
- 變數名稱 : car\_price, truck\_price
- 計算時間 : 0.029秒
- 箱型圖 :



Part 6.

# 資料的離度：變動性

# 心臟沒這麼大顆

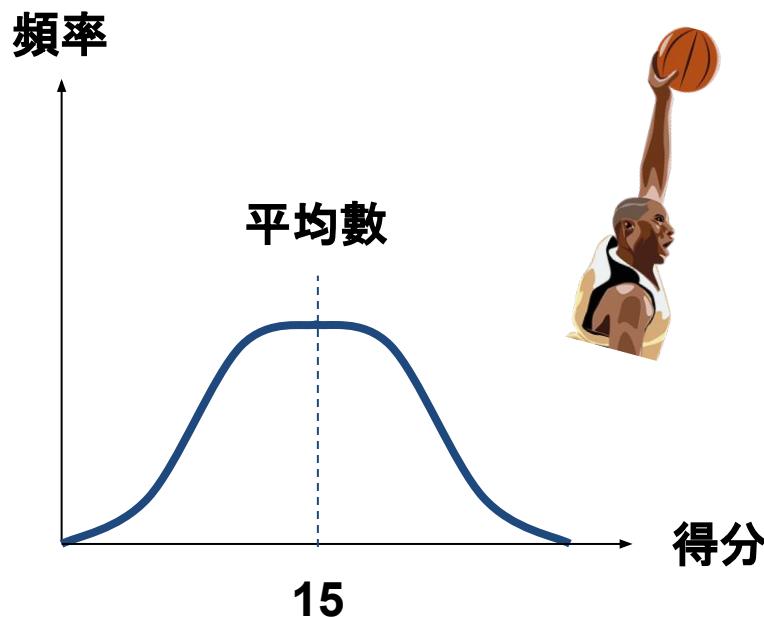


「我想要穩定賺錢就好，  
不想要高風險，可以嗎？」

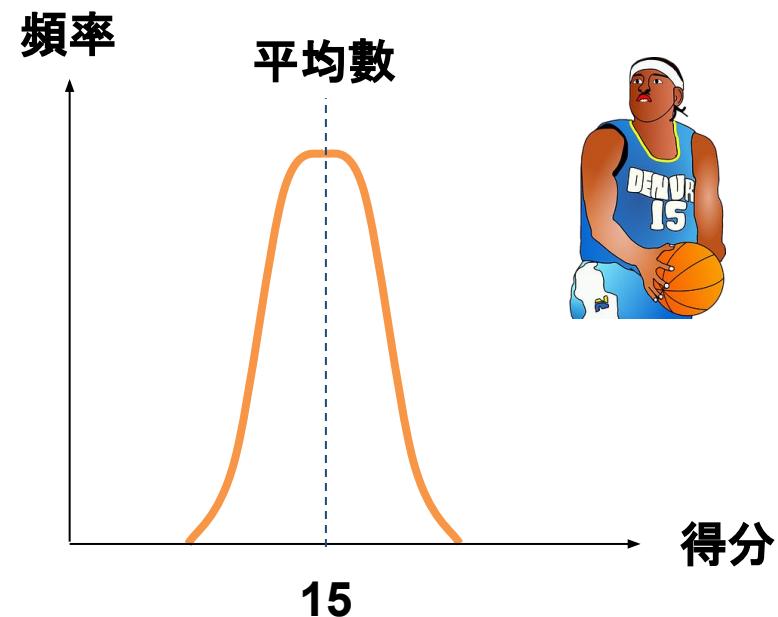


# 教練也想知道該派誰好？

球員阿布的得分

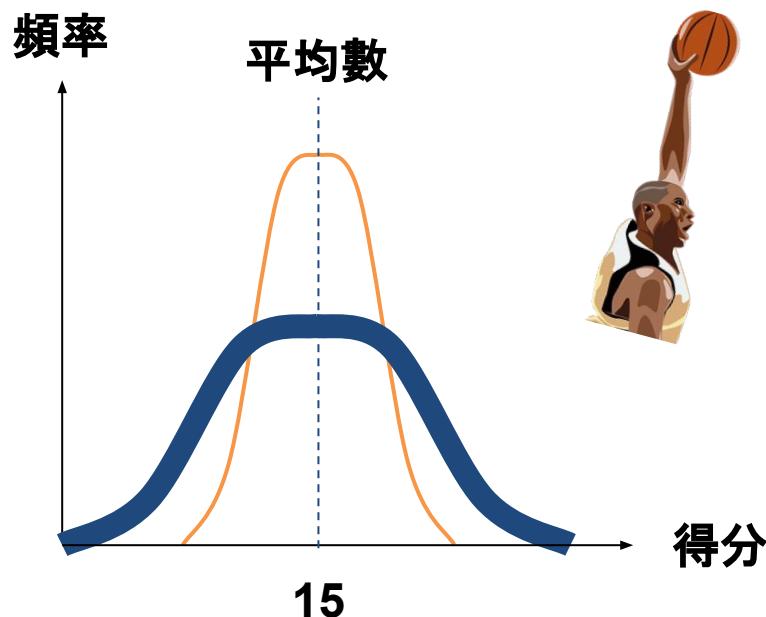


球員阿丁的得分



「那個球員比較能預測他下一場的表現呢？」

# 球員阿布的得分

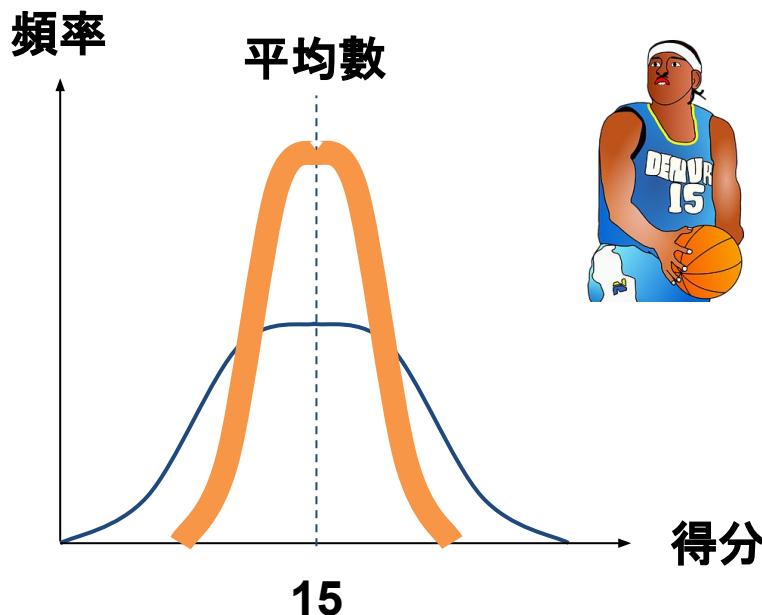


- 資料比較分散
- 離平均數比較遠
- 狀況好時會得高分
- 狀況不好時可能很糟

不穩定  
風險太大



# 球員阿丁的得分



- 資料比較集中
- 離平均數比較近
- 容易預測每一場的可能得分

很穩定  
風險小



# 描述離度的另一種方法

資料  
分散性

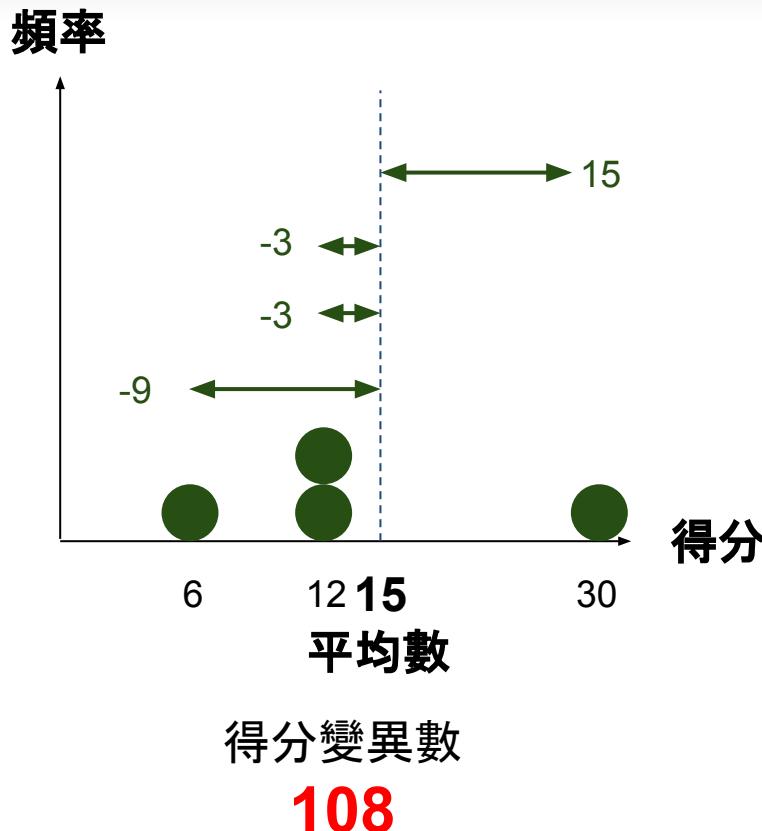


五數綜合  
(全距、四分位數距)

資料  
變動性

每個值到平均數的平均距離  
(變異數、標準差)

# 資料變動性的第一招 變異數

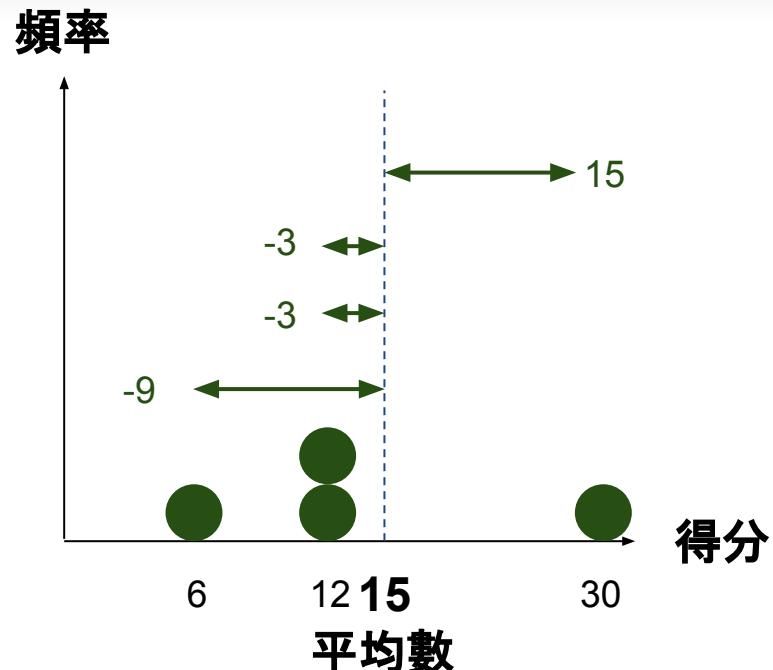


變異數是呈現所有數值與平均數的平均概念

- 計算每一個數值到標準差的距離=離均差
- **離均差平方 (避免負數)**
- 加總
- 除與資料總數-1  
(樣本資料的貝索校正)



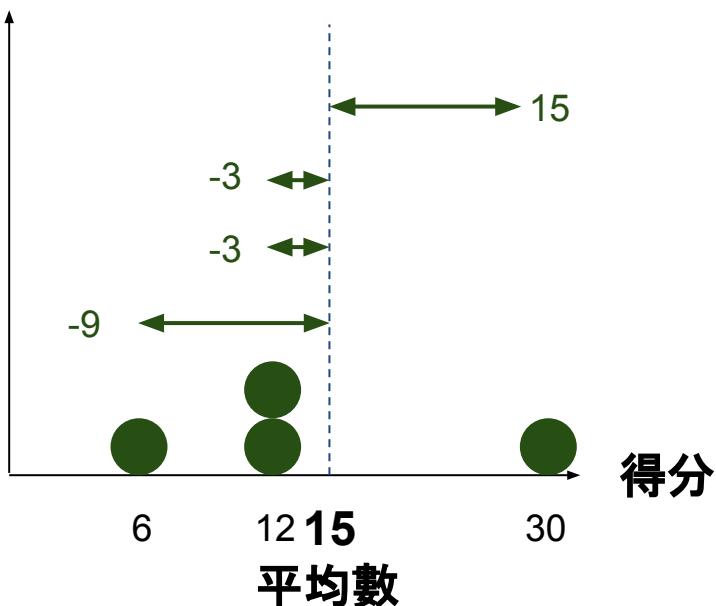
# 變異數的限制



哪裡像距離平均？

# 資料變動性的第二招 標準差

頻率



得分變異數

108

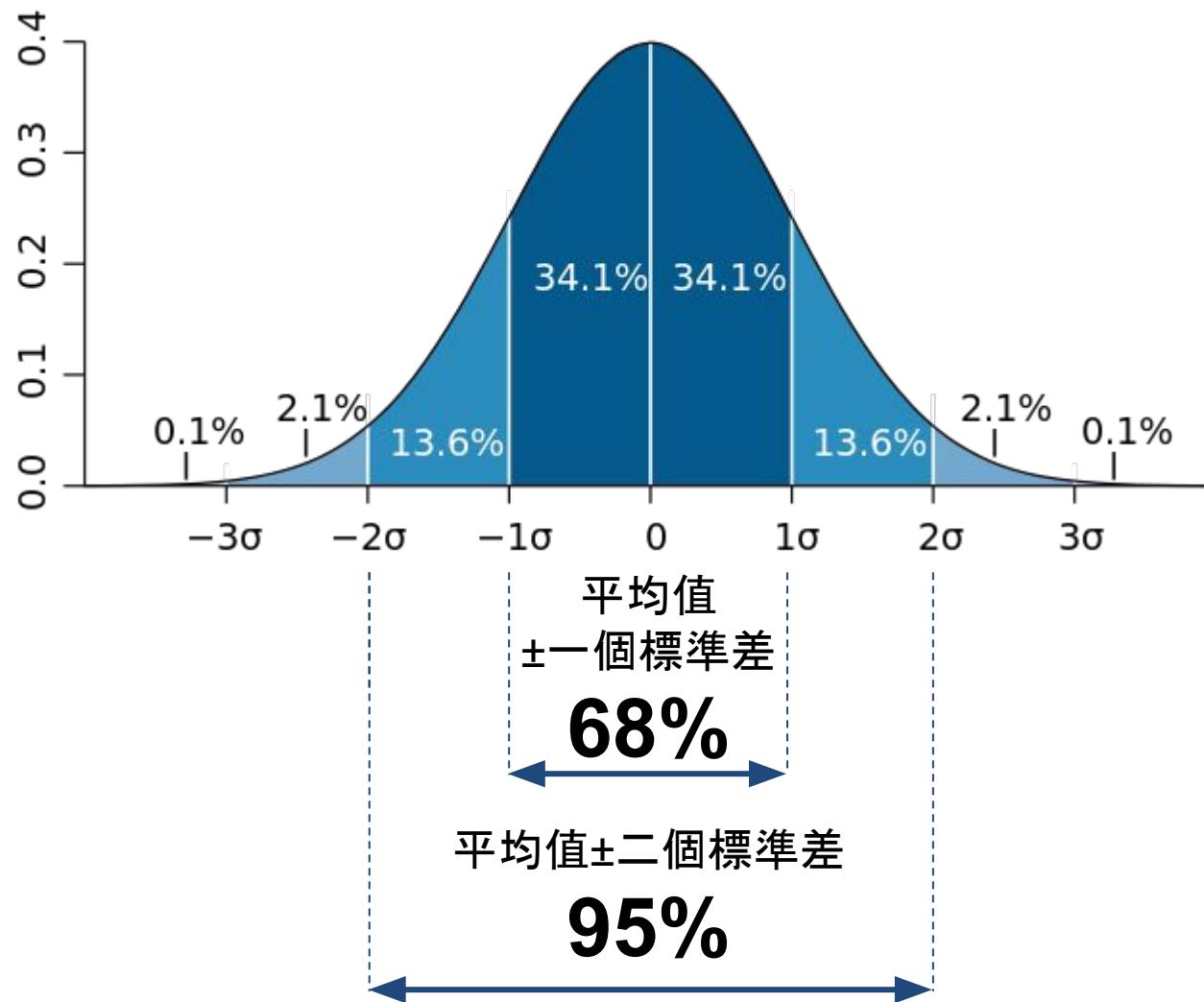
取平方根  $\sqrt{ }$

得分標準差

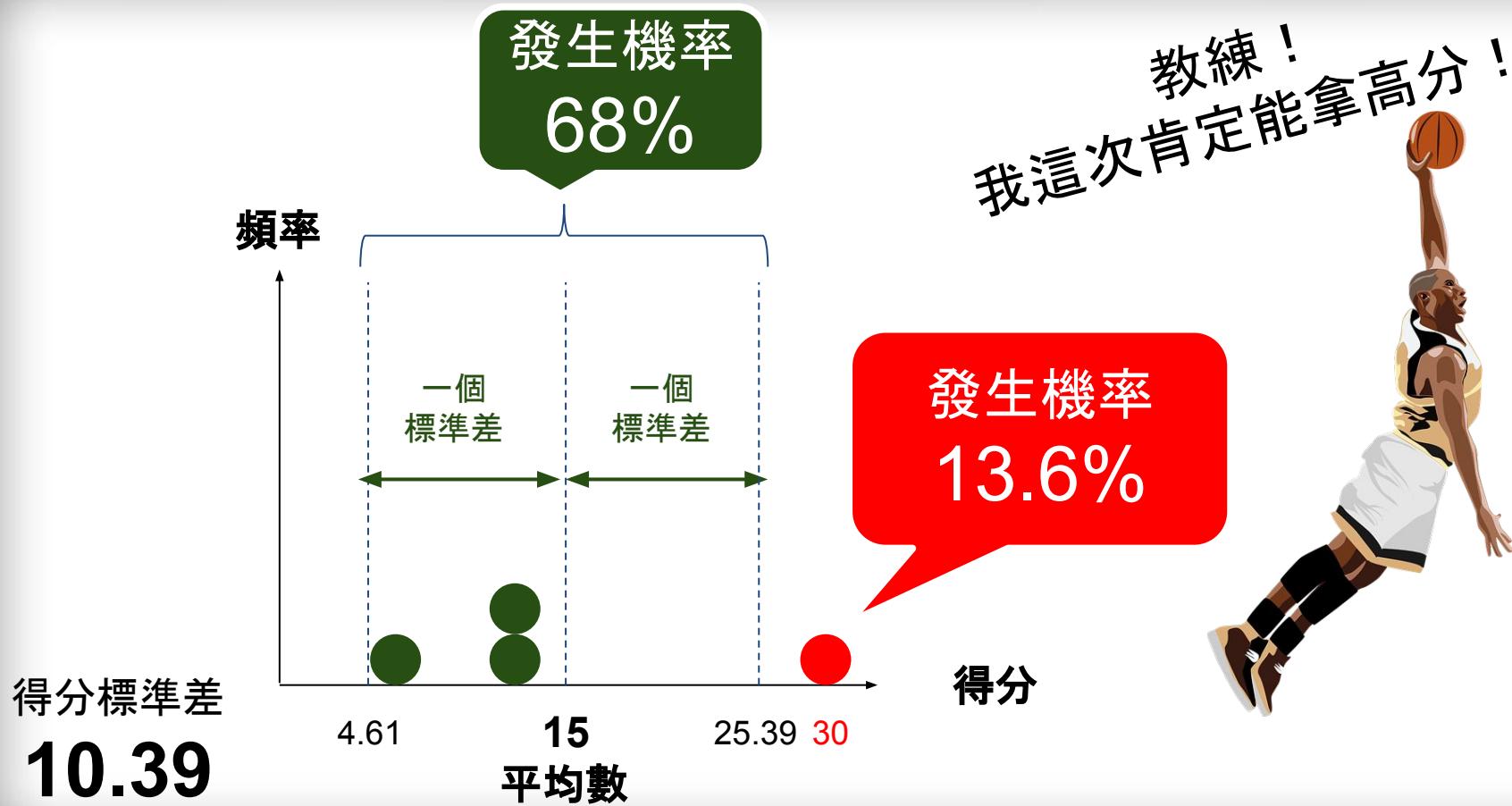
10.39

很像  
距離平均

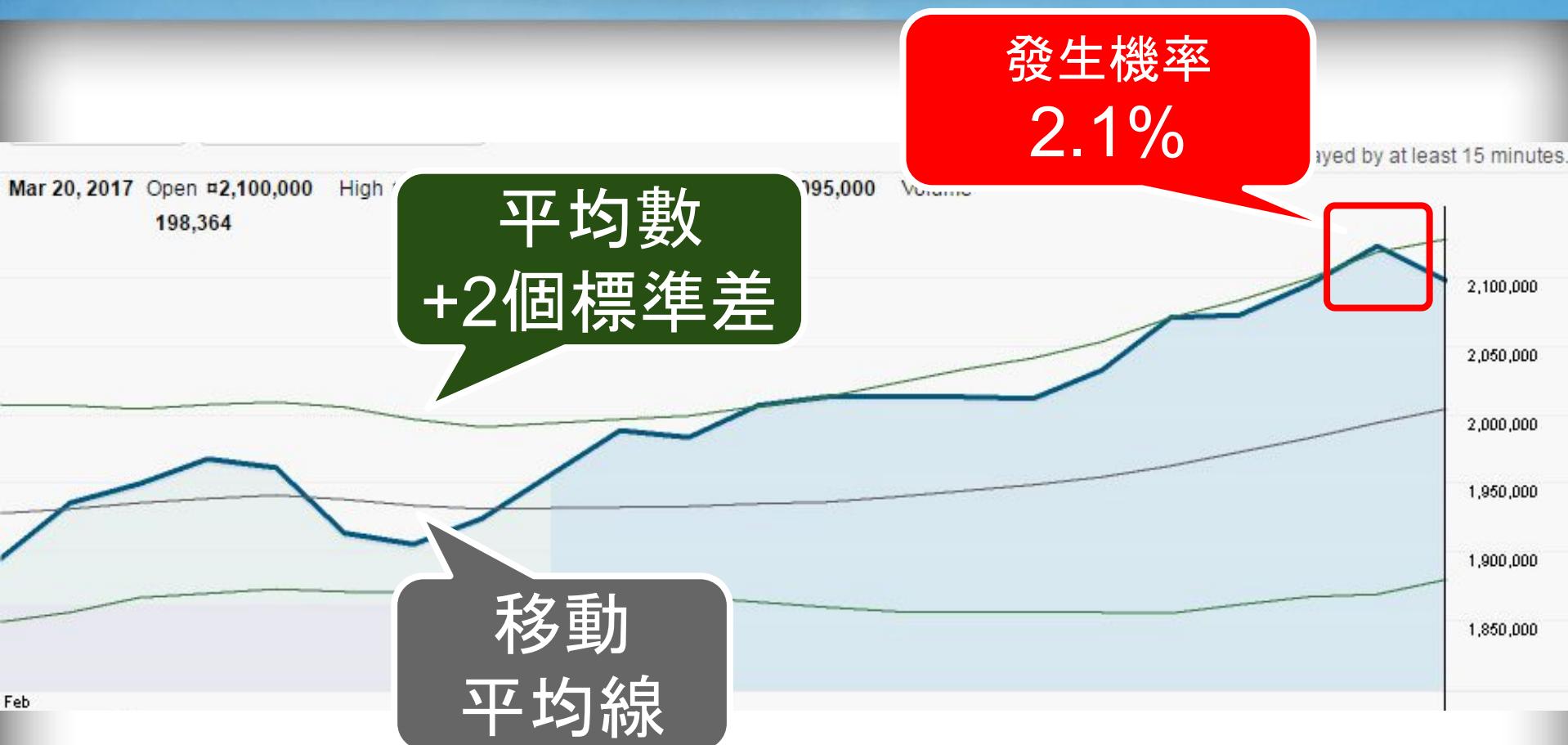
# 常態分配的規則 用標準差預測發生機率



# 得分30的機率？



# 股票的標準差 布林線 B-Band



# 母體與樣本的算法不同

母體標準差

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$



=STDEV $\mathbf{P}$ ()

樣本標準差

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

貝索校正

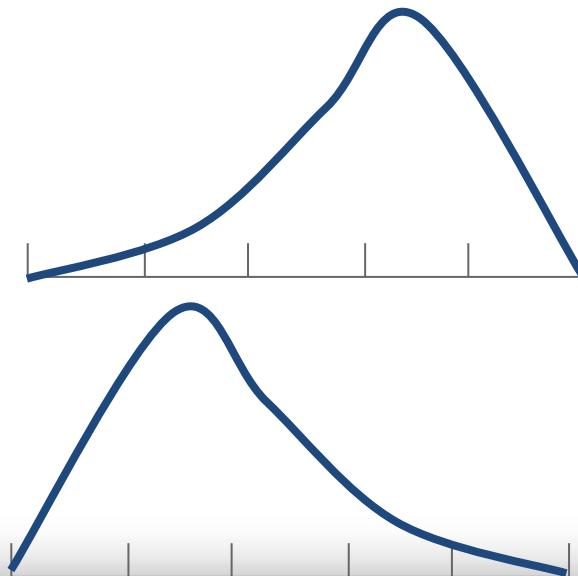


=STDEV()

# 如何選擇適當的綜合數值描述？

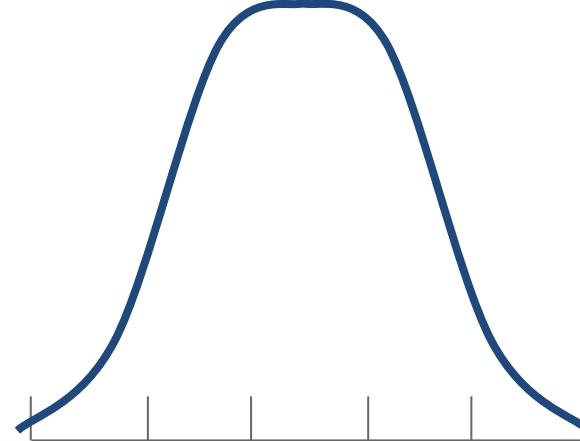
- 資料有偏斜分佈、長尾巴
- 資料有離群值

選擇「五數綜合」



- 資料分佈對稱(常態分配)
- 沒有離群值

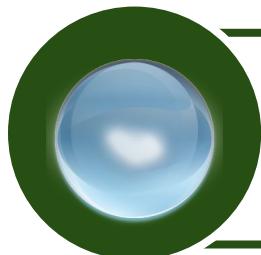
選擇「平均數 & 標準差」



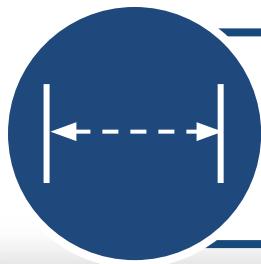
# 資料敘述級的步驟



**統計圖表 (全貌概況)**



**均值 (代表數值)**  
平均數、中位數、眾數



**離度 (變化數值)**

分散性(五數綜合)、變動性(變異數、標準差)

Part 7.

# 實作：變異數與標準差的計算

# 哪一種車價格變動大？



「我只想買進貨價格穩定的車子種類。」

汽車與卡車...

註解 共用 說明

所有變更都已儲存到雲端硬碟

更多選項

fx 汽車價格(千美元)

	A	B
1	汽車價格(千美元)	卡車價格(千美元)
2	22	46
3	28	19
4	42	21
5	24	19
6	34	17
7	62	26
8	27	20
9	33	32
10	39	21
11	22	36
12	25	12
13	32	27
14	28	21
15	40	26
16	44	26
17	40	14

# 汽車跟卡車的價格

汽車案例數量: 100筆

卡車案例筆數: 40筆



# W05-d.

## 變異數與標準差的計算

1. 插入汽車變異數的公式與資料範圍
2. 插入汽車標準差的公式與資料範圍
3. 插入卡車的變異數與標準差  
的公式與資料範圍



# 1. 插入變異數的公式與資料範圍

1. 點選「變異數」旁邊的空格 (E17)
2. 點選「公式列」的空白處
3. 輸入公式：

=VAR(A2:A101)

2

1

	A	B	C	D	E
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與離度	
13		32	27	上四分位數Q3	
14		28	21	下四分位數Q1	
15		40	26	內四分位距	
16		44	26		
17		40	14	變異數	
		31	22	標準差	
13		27			
17		60			
21		19	35		
22		19	43		
23		24	32		
24		46	23		
25		14	28		
--		--	--		

+ data template demo figure

## 2. 插入標準差的公式與資料範圍

1. 點選「標準差」旁邊的空格 (E18)
2. 點選「公式列」的空白處
3. 輸入公式：

=STDEV(A2:A101)

2

1

	A	B	C	D	E
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與離度	
13	32	27		上四分位數Q3	
14	28	21		下四分位數Q1	
15	40	26		內四分位距	
16	44	26			
17	40	14		變異數	
	31	22		標準差	
13	27				
17	60				
21	19	35			
22	19	43			
23	24	32			
24	46	23			
25	14	28			
--	--	--			

data template demo figure

### 3. 插入卡車的 變異數與標準差 的公式與資料範圍

- 卡車變異數  
H17

$=\text{VAR}(\text{B2:B41})$

- 卡車標準差  
H18

$=\text{STDEV}(\text{B2:B41})$

汽車與卡車的價格-公式用			
	B	C	D
E)		汽車的中心與離度	卡車的中心與離度
2			
7		上四分位數Q3	上四分位數Q3
14	21	下四分位數Q1	下四分位數Q1
15	26	內四分位距	內四分位距
16	26		
17	14	變異數	變異數
18	22	標準差	標準差
19	27		
20	60		
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			
101			
102			
103			
104			
105			
106			
107			
108			
109			
110			
111			
112			
113			
114			
115			
116			
117			
118			
119			
120			
121			
122			
123			
124			
125			
126			
127			
128			
129			
130			
131			
132			
133			
134			
135			
136			
137			
138			
139			
140			
141			
142			
143			
144			
145			
146			
147			
148			
149			
150			
151			
152			
153			
154			
155			
156			
157			
158			
159			
160			
161			
162			
163			
164			
165			
166			
167			
168			
169			
170			
171			
172			
173			
174			
175			
176			
177			
178			
179			
180			
181			
182			
183			
184			
185			
186			
187			
188			
189			
190			
191			
192			
193			
194			
195			
196			
197			
198			
199			
200			
201			
202			
203			
204			
205			
206			
207			
208			
209			
210			
211			
212			
213			
214			
215			
216			
217			
218			
219			
220			
221			
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			
234			
235			
236			
237			
238			
239			
240			
241			
242			
243			
244			
245			
246			
247			
248			
249			
250			
251			
252			
253			
254			
255			
256			
257			
258			
259			
260			
261			
262			
263			
264			
265			
266			
267			
268			
269			
270			
271			
272			
273			
274			
275			
276			
277			
278			
279			
280			
281			
282			
283			
284			
285			
286			
287			
288			
289			
290			
291			
292			
293			
294			
295			
296			
297			
298			
299			
300			
301			
302			
303			
304			
305			
306			
307			
308			
309			
310			
311			
312			
313			
314			
315			
316			
317			
318			
319			
320			
321			
322			
323			
324			
325			
326			
327			
328			
329			
330			
331			
332			
333			
334			
335			
336			
337			
338			
339			
340			
341			
342			
343			
344			
345			
346			
347			
348			
349			
350			
351			
352			
353			
354			
355			
356			
357			
358			
359			
360			
361			
362			
363			
364			
365			
366			
367			
368			
369			
370			
371			
372			
373			
374			
375			
376			
377			
378			
379			
380			
381			
382			
383			
384			
385			
386			
387			
388			
389			
390			
391			
392			
393			
394			
395			
396			
397			
398			
399			
400			
401			
402			
403			
404			
405			
406			
407			
408			
409			
410			
411			
412			
413			
414			
415			
416			
417			
418			
419			
420			
421			
422			
423			
424			
425			
426			
427			
428			
429			
430			
431			
432			
433			
434			
435			
436			
437			
438			
439			
440			
441			
442			
443			
444			
445			
446			
447			
448			
449			
450			
451			
452			
453			
454			
455			
456			
457			
458			
459			
460			
461			
462			
463			
464			
465			
466			
467			
468			
469			
470			
471			
472			
473			
474			
475			
476			
477			
478			
479			
480			
481			
482			
483			
484			
485			
486			
487			
488			
489			
490			
491			
492			

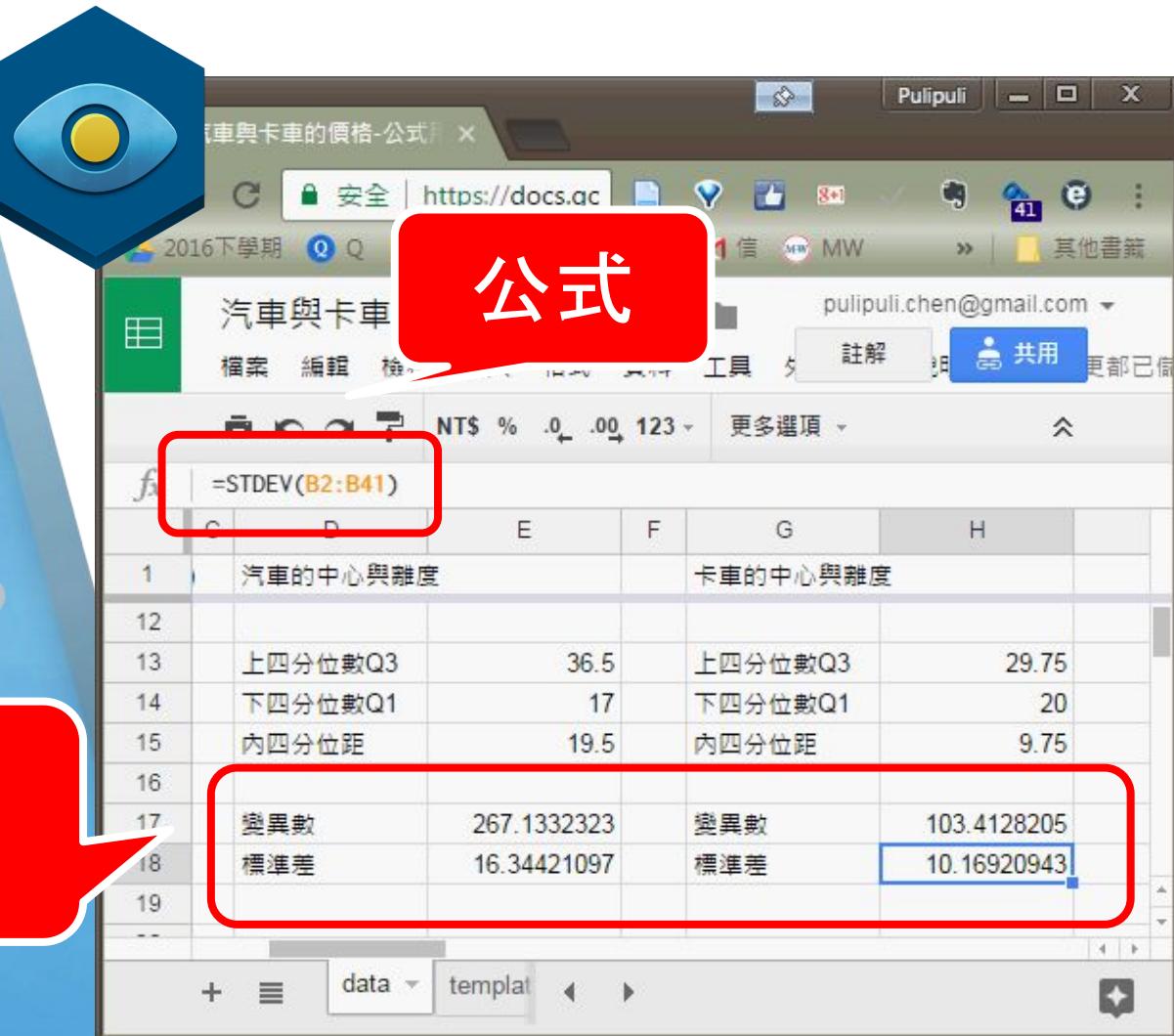
# W05-d.

## 變異數與標準差的計算

上機實作

變異數  
標準差

公式



汽車與卡車的價格-公式

2016下學期

汽車與卡車

檔案 編輯 檢視 工具 註解 共用

	汽車的中心與離度	卡車的中心與離度
13	上四分位數Q3	36.5
14	下四分位數Q1	17
15	內四分位距	19.5
17	變異數	267.1332323
18	標準差	16.34421097
19		變異數
		103.4128205
		標準差
		10.16920943

Part 8.

# 練習： 不同年齡對網路郵局的滿意度

# 104年度 不同年齡民眾對 於 網路郵局滿意 度

20-未滿30歲:45筆

30-未滿40歲:51筆

40-未滿50歲:47筆

50-未滿60歲:47筆

60-未滿65歲:16筆

65歲及以上:34筆

104年度 不同年齡民眾對於 網路郵局滿意度						
	A	B	C	D	E	F
1	20-未滿30歲	30-未滿40歲	40-未滿50歲	50-未滿60歲	60-未滿65歲	65歲及以上
2	1	2	3	2	2	2
3	1	2	4	2	2	2
4	2	2	2	1	2	1
5	1	2	2	2	2	1
6	1	2	2	2	2	1
7	2	2	2	2	2	2
8	3	2	2	2	3	2
9	1	1	1	1	2	2
10	2	2	1	2	2	2
11	2	2	2	2	1	2
12	1	1	1	2	2	2
13	1	2	2	2	1	1
14	2	1	2			
15	1	2	2			
16	2	2	1			
17	2	2	2			
18	2	2	2			
19	2	2	1			
20	2	2	1			
21	2	2	1			
22	2	1	2			
23	2	2	2			
24	2	2	2			

# W05-e. 不同年齡對 網路郵局的滿意度

計算不同年齡層的均值與離度

- 均值: 平均數、中位數、眾數
- 五數綜合:  
最大值、最小值、全距  
上四分位數Q3、下四分位數Q1、  
內四分位距
- 變動性: 變異數、標準差



# W05-e. 不同年齡對 網路郵局的滿意度

課後作業



104年度不同年齡民眾對於網路郵局滿意度

連結共用設定已「開啟」。復原

=STDEV(F2:F35)

	G	H	I	J	K	L	M	N
1		20-未滿30歲	30-未滿40歲	40-未滿50歲	50-未滿60歲	60-未滿65歲	65歲及以上	
9	最大值	3	4	4	2	3	3	3
10	最小值	1	1	1	1	1	1	1
11	全距	2	3	3	1	2	2	2
12								
13	上四分位數Q3	2	2	2	2	2	2	2
14	下四分位數Q1	1	2	2	1	2	1	1
15	內四分位距	1	0	0	1	0	1	1
16								
17	變異數	0.3131313131	0.2949019608	0.3182238668	0.235892692	0.1958333333	0.2745098039	
18	標準差	0.5595813731	0.5430487646	0.5641133457	0.4856878544	0.4425306016	0.523936832	
19								
20								

*Thank you for  
your attention*

