

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



狂



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





# 汽車跟卡車的價格

汽車案例數量:100筆

卡車案例筆數:40筆

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

更多選項

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			
15	40			
16	44			
17	40			

+ data



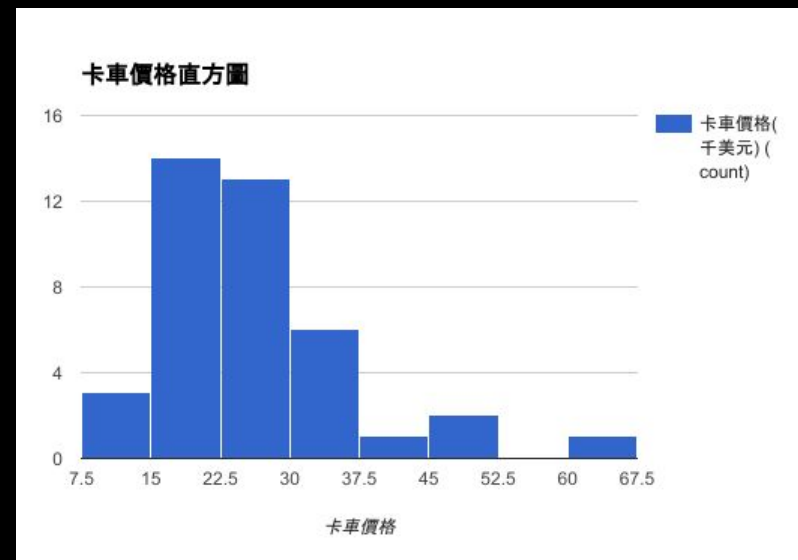
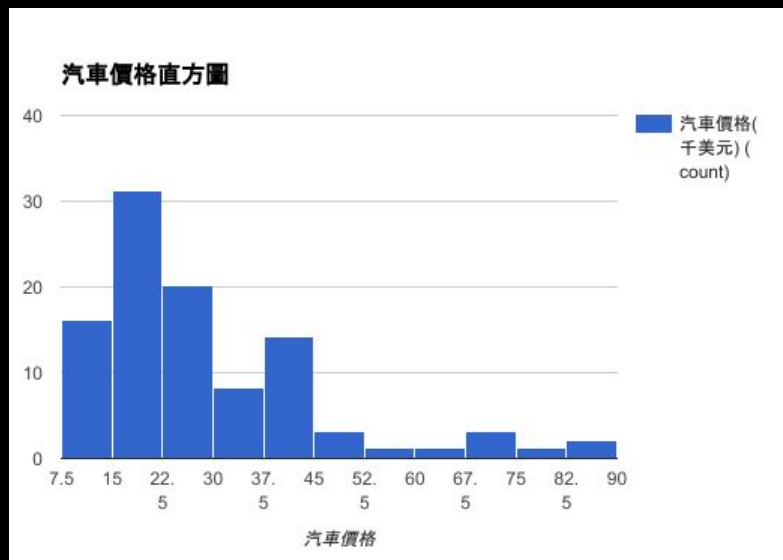
# 本週的老闆



「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

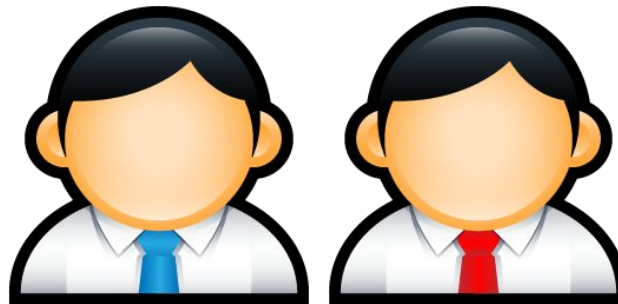


Part 1.

# 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



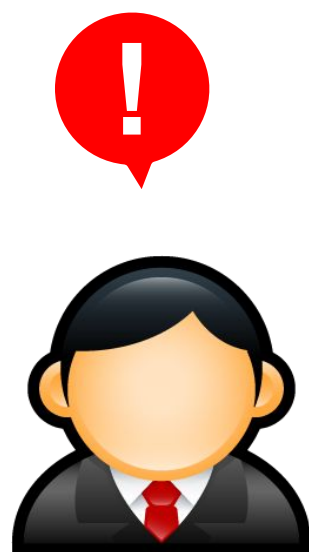
員工  
\$500



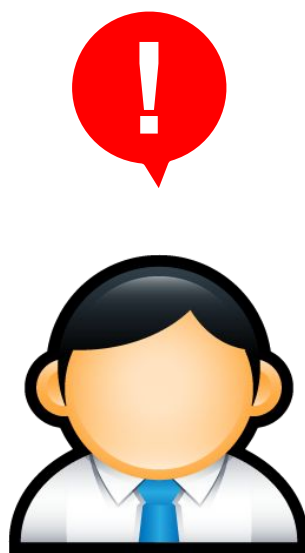
員工  
\$500



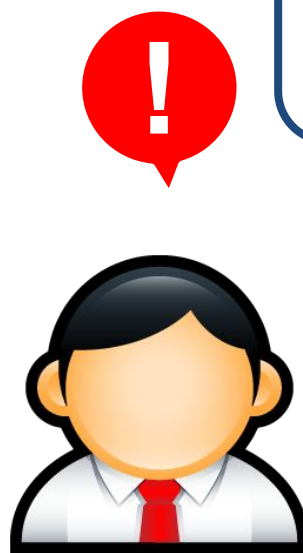
# 黑心企業？



經理  
\$4500



員工  
\$500



員工  
\$500

週薪均值  
**500美元**而已！  
回去工作！



CEO  
\$50500

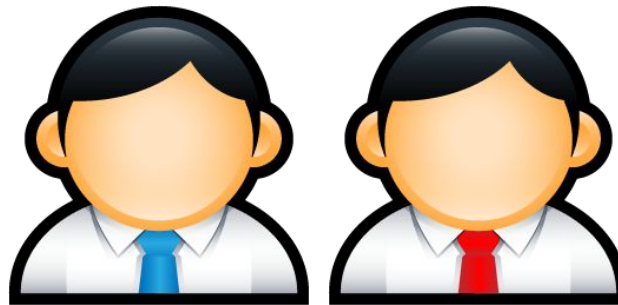
# 誰在說謊？

週薪均值  
**14000美元**



經理  
\$4500

週薪均值  
**2500美元**



員工  
\$500

員工  
\$500

週薪均值  
**500美元**



CEO  
\$50500

# 誰在說謊？

平均數

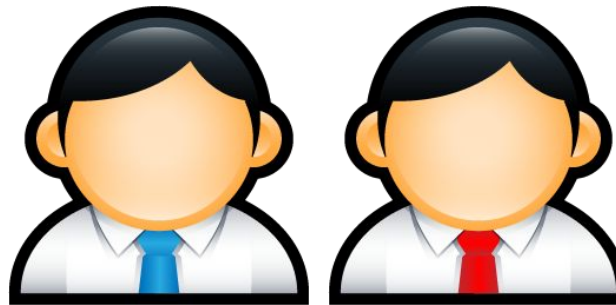
週薪均值  
**14000美元**



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**

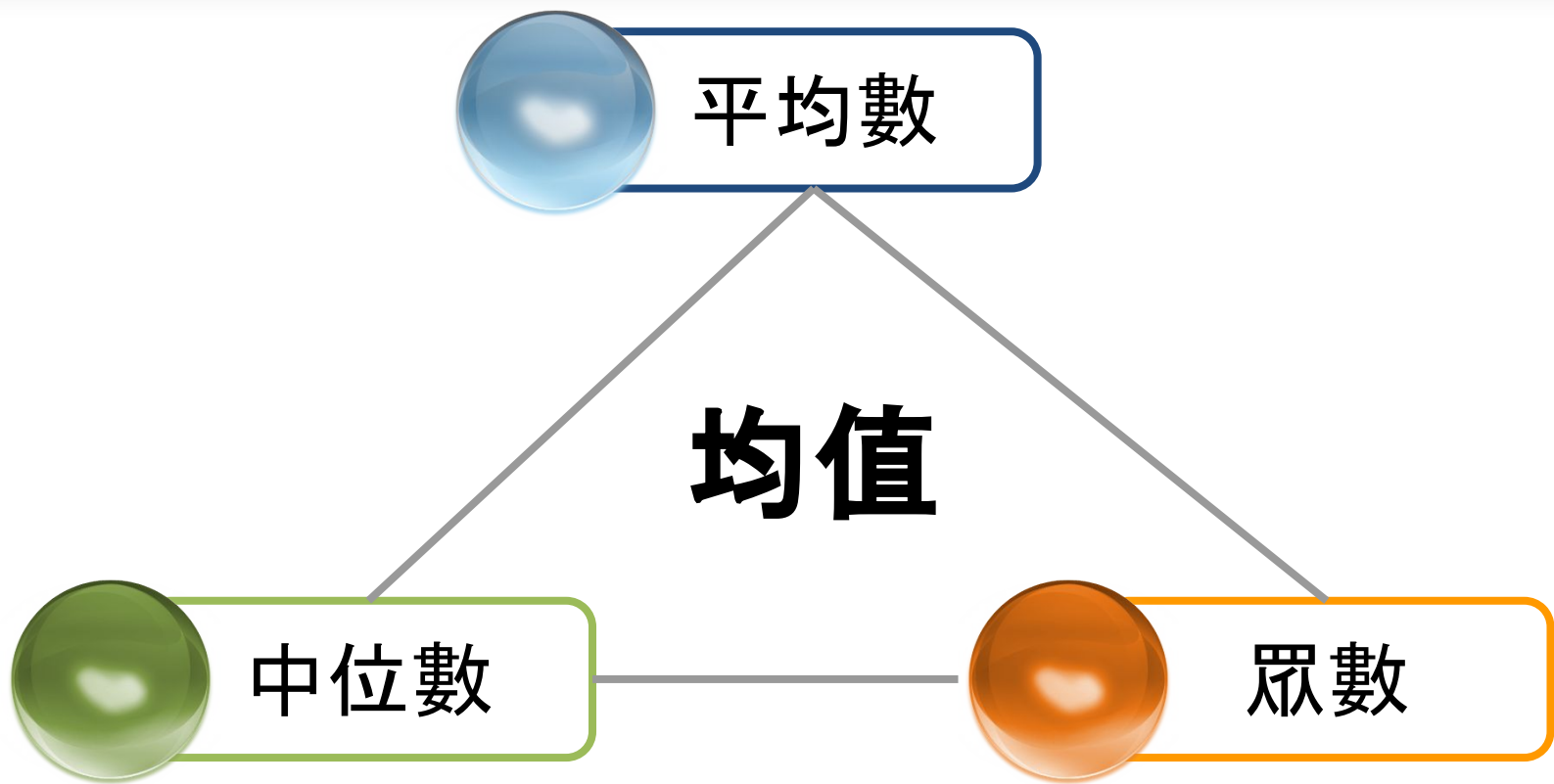


CEO  
\$50500

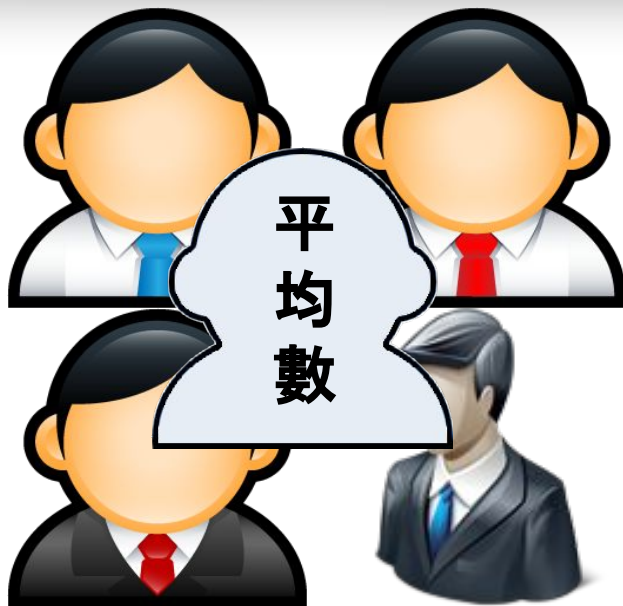


# 資料中心的三神器

## 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			

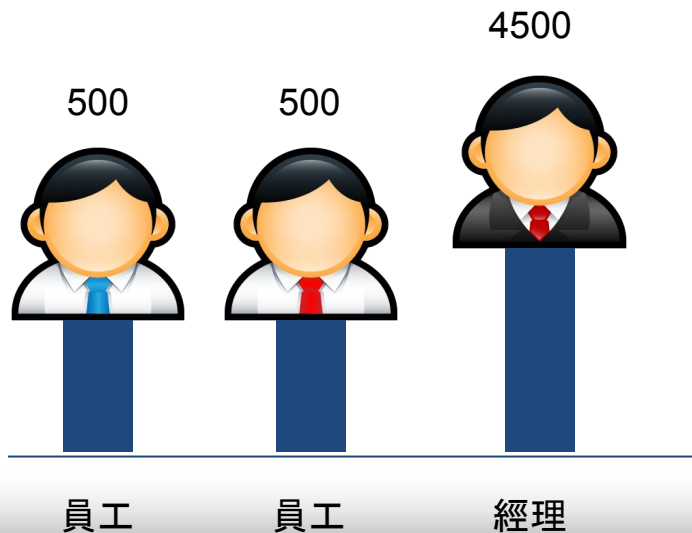
平均數=34.6



# 平均數的問題

離群值

平均數：  
1833.33

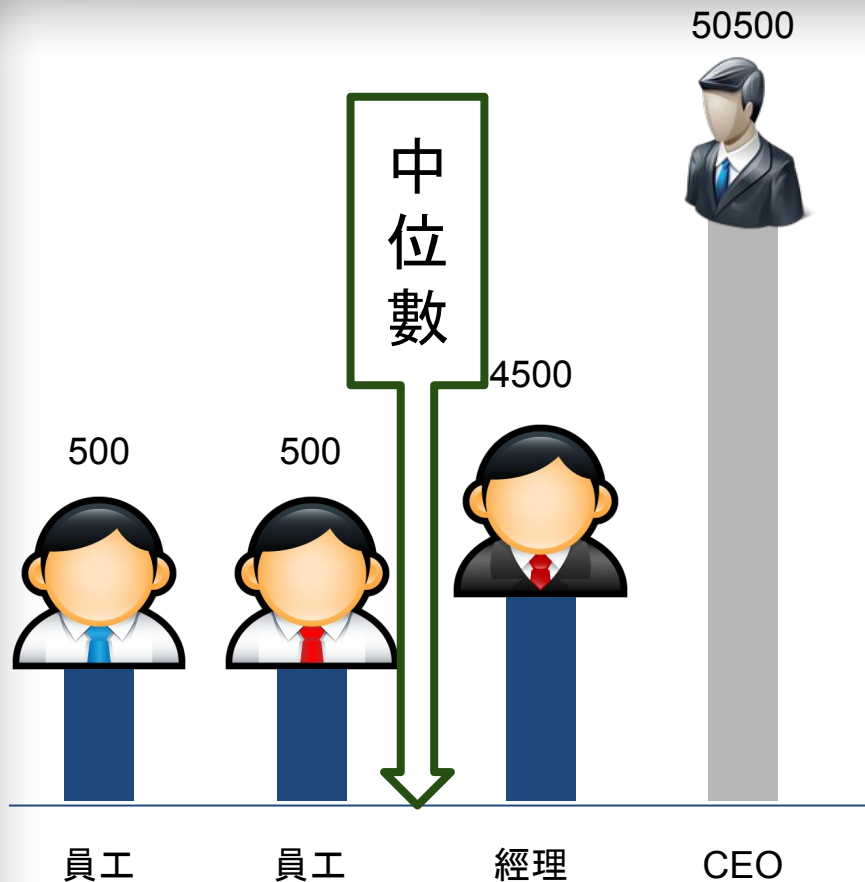


平均數：  
**14000**





# 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**



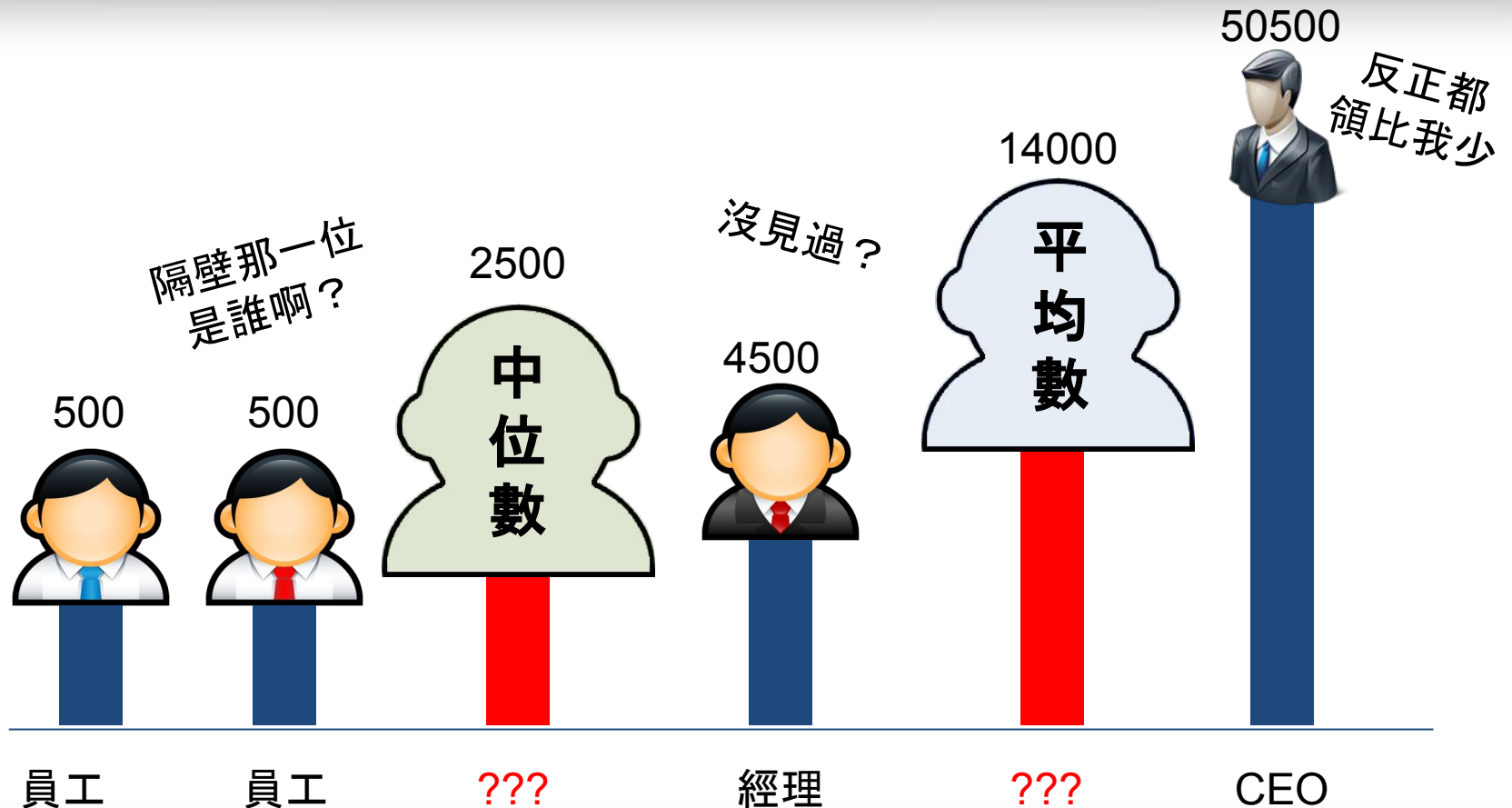
**=MEDIAN()**

# 中位數的使用場合

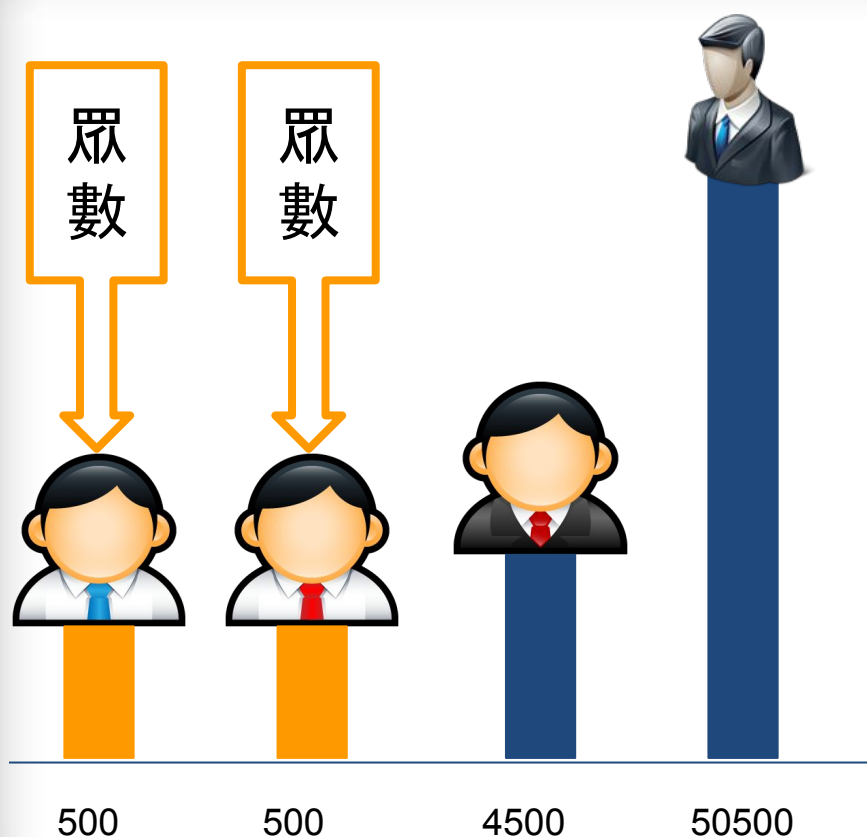
## 工業及服務業受僱員工每人每月薪資及經常性薪資 (單位:新臺幣元)

	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

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

週薪眾數=

500出現最多次(2次)=

500

=MODE()

# 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

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



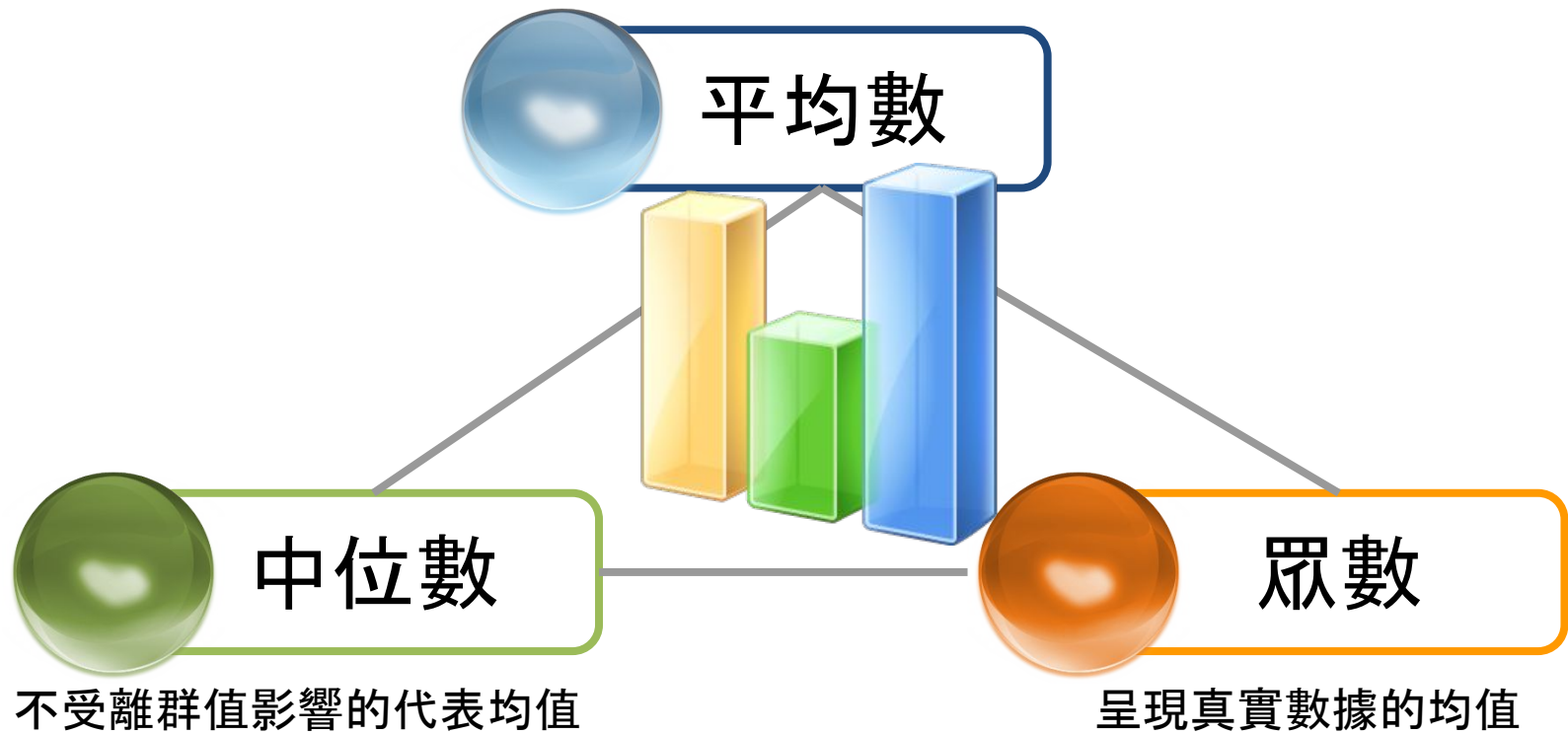
小丑糖
蜂糖
幼兒糖
蜂糖
幼兒糖
蜂糖
蜂糖
小丑糖
幼兒糖
蜂糖



眾數  
**蜂糖**

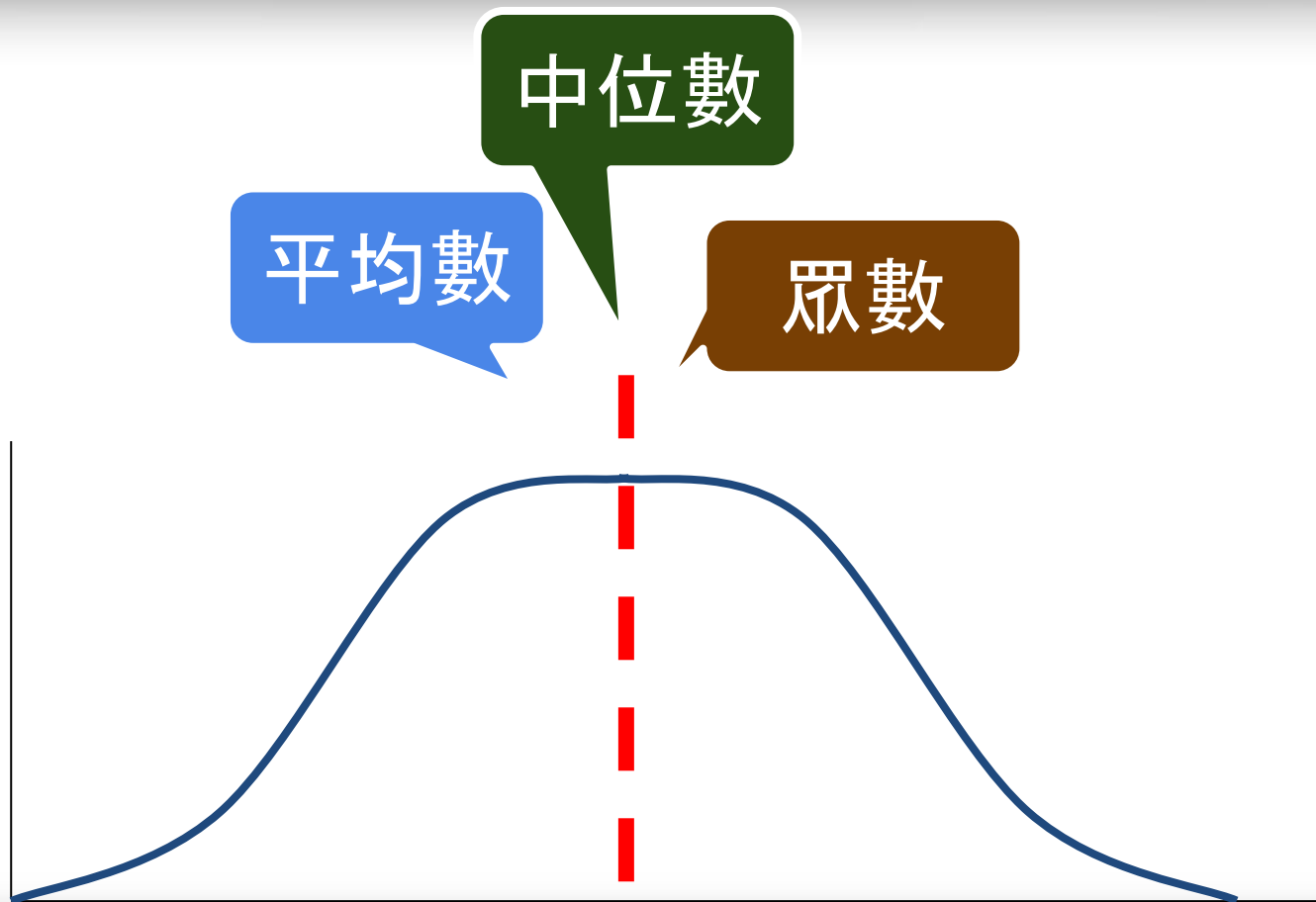
# 均值與資料形狀的關係

最常見的抽象均值



# 均值與資料形狀的關係

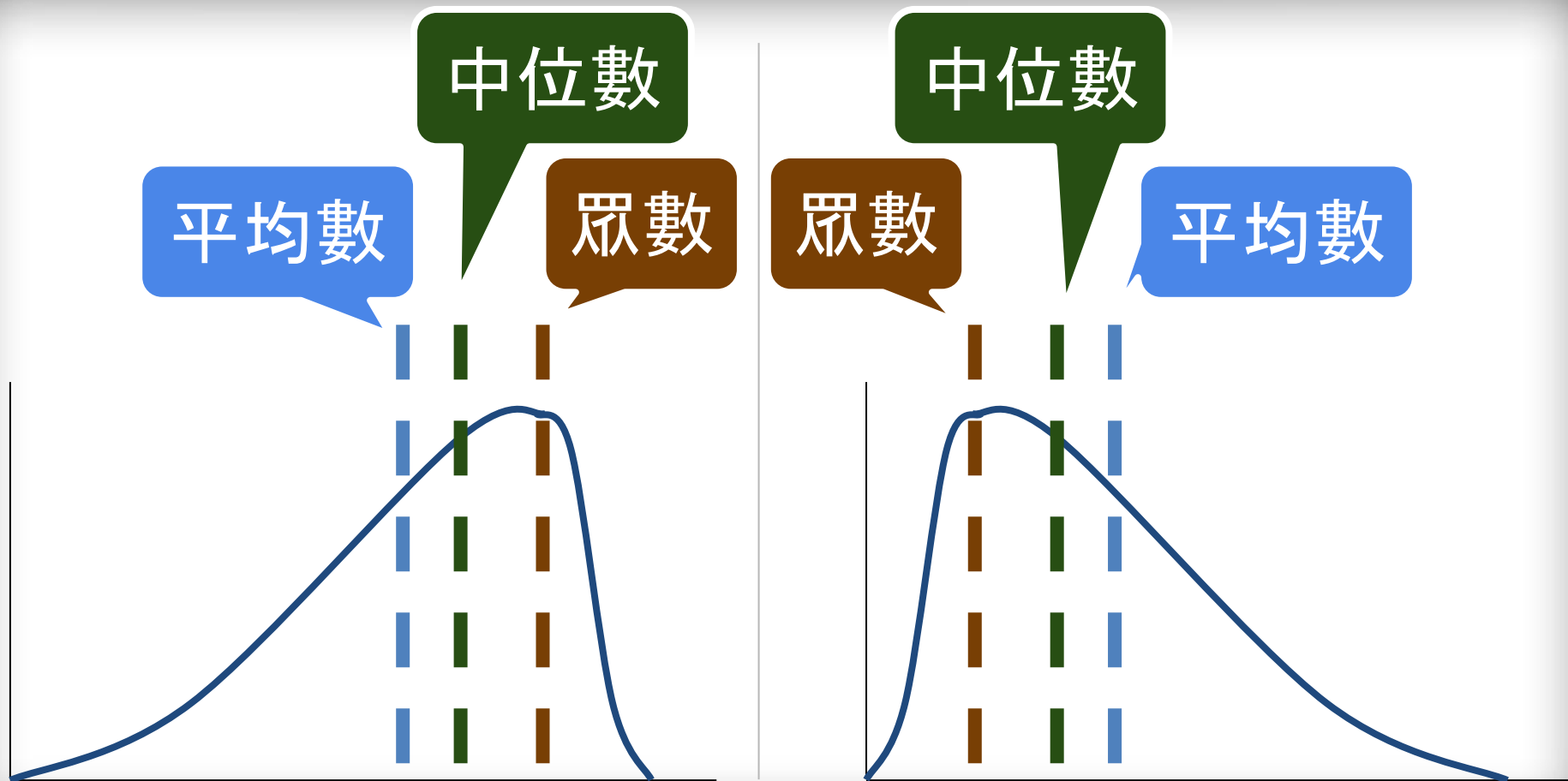
## 對稱=常態分配



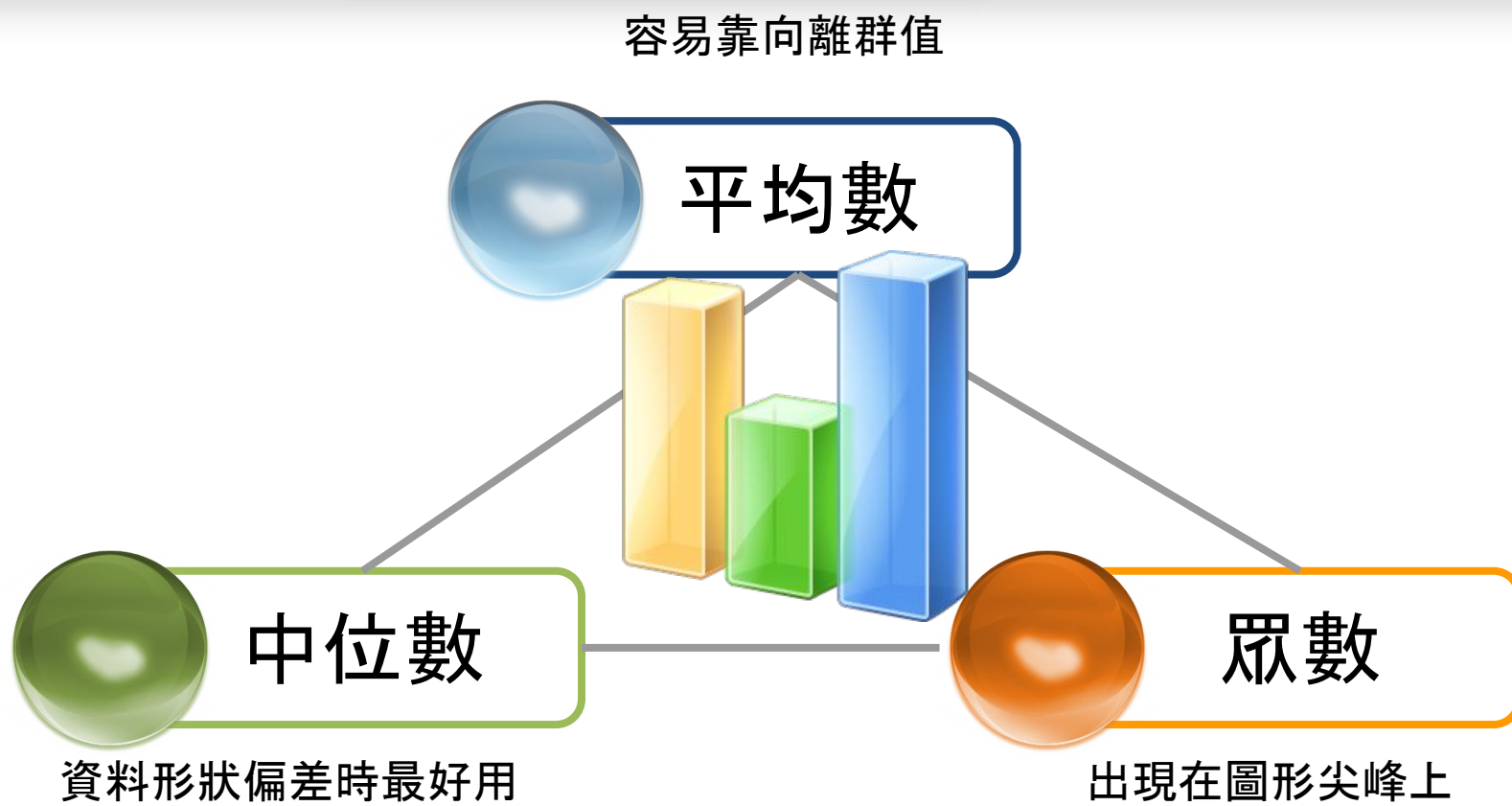
# 均值與資料形狀的關係

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

更多選項

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	14		

data

# 汽車的價格

汽車案例數量:100筆



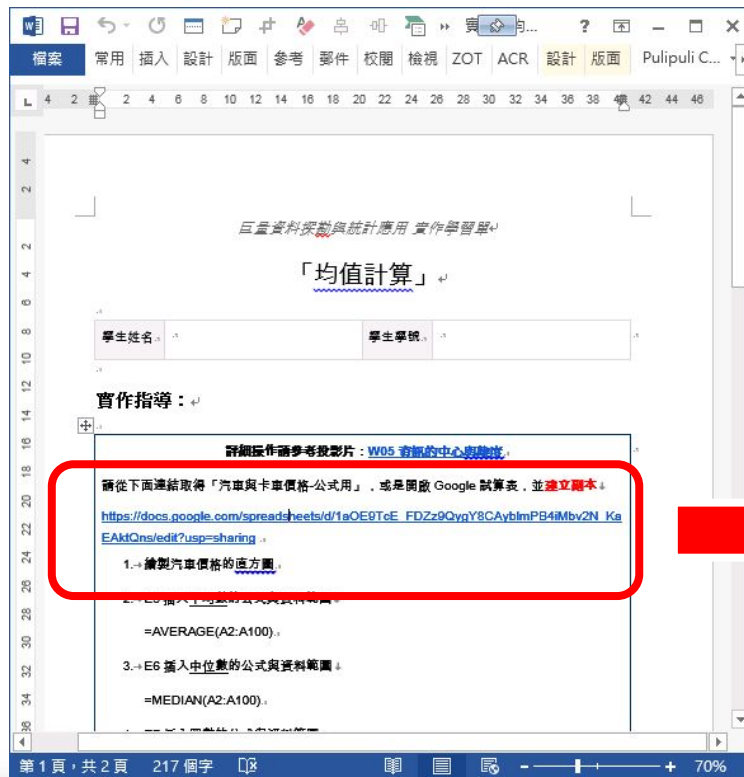
# W05-a. 均值計算

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



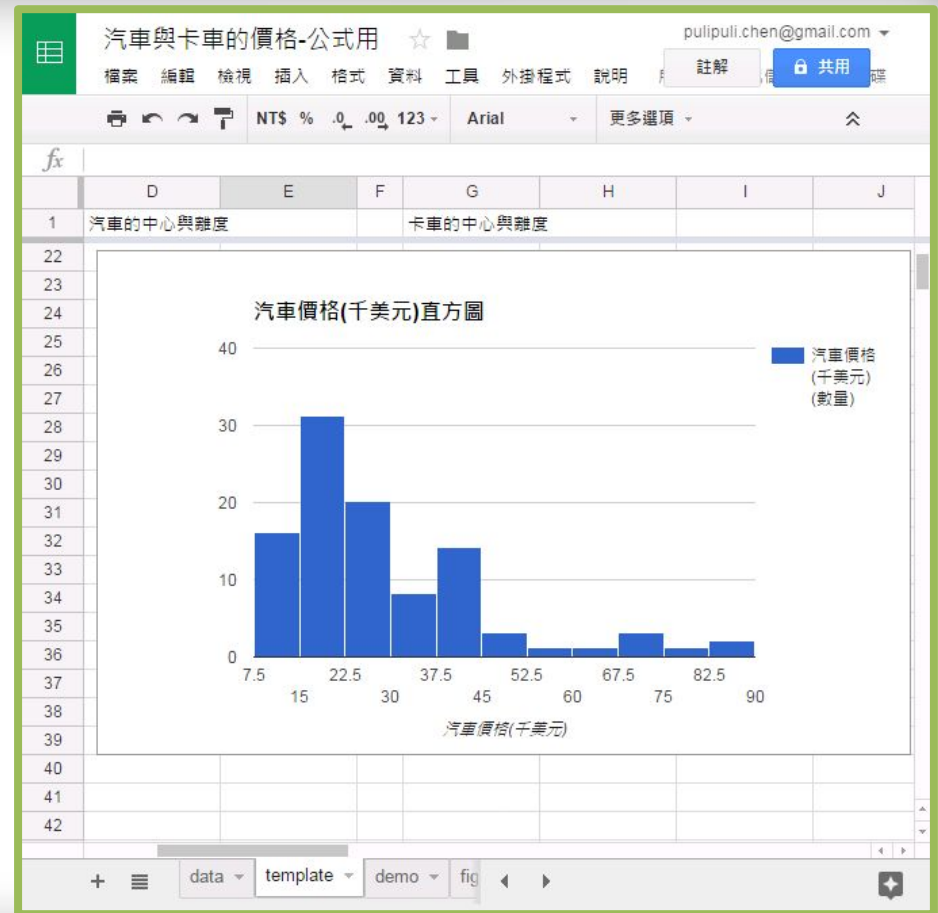
汽車與卡車的價格-公式用

	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	
	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	



# 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」旁邊的空白處
2. 點選「公式列」
3. 輸入公式：

**=AVERAGE(A2:A101)**

這是  
公式

公式

資料範圍

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與難度		
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		
	27	20				
	33	32		最大值		
	39	21		最小值		
	22	36		全距		
	25	12				
	32	27		上四分位數Q3		
	28	21		下四分位數Q1		

The formula bar is highlighted with a red circle and the number 2. A cell in the table (E5) is highlighted with a red circle and the number 1.

# 資料範圍：單格

A:欄名

2:列名

資料範圍:A2

值:21.5

2:列名

A:欄名

汽車與卡車的價格-公式用		
檔案 編輯 檢視 插入 格式		
	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：

	檔案	編輯	檢視	插入	格式
fx	22	NT\$	%	.0	.0
	A				
1	汽車價格(千美元)				
2	22	46			
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)

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與難度		
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		
	27	20				
	33	32		最大值		
	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		

The formula bar is highlighted with a red circle and the number 2. The cell E6 is highlighted with a red circle and the number 1.



## 4. 插入眾數的公式與資料範圍

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

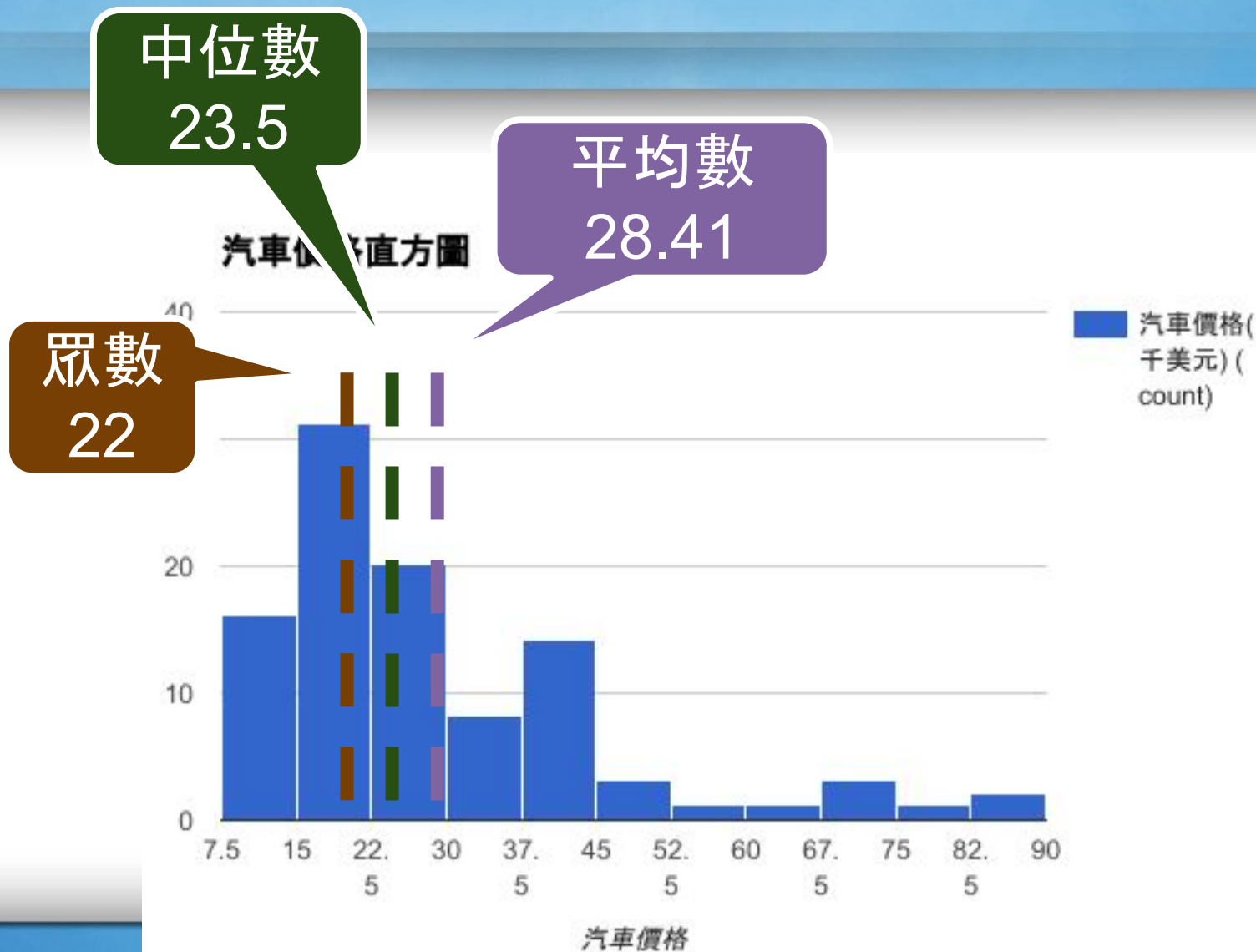
**=MODE(A2:A101)**

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	汽車價格(千美元)	卡車價格(千美元)		汽車的中心與難度		
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		
	27	20				
	33	32		最大值		
	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		

The formula bar is highlighted with a red circle and the number 2. Cell E7 is highlighted with a red circle and the number 1.

# 觀察均值在資料分佈的位置



# W05-a. 均值計算

上機實作



公式

=MODE(A2:A101)

均值

汽車與卡車的價格-公式			
檔案 編輯 檢視 插入			
NT\$ 0%			
fx			
=MODE(A2:A101)			
	A	B	C
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
4	42	21	A101
5	24	19	平均數
6	34	17	中位數
7	62	26	眾數
8	27	20	

28.41

23.5

22

Part 3.

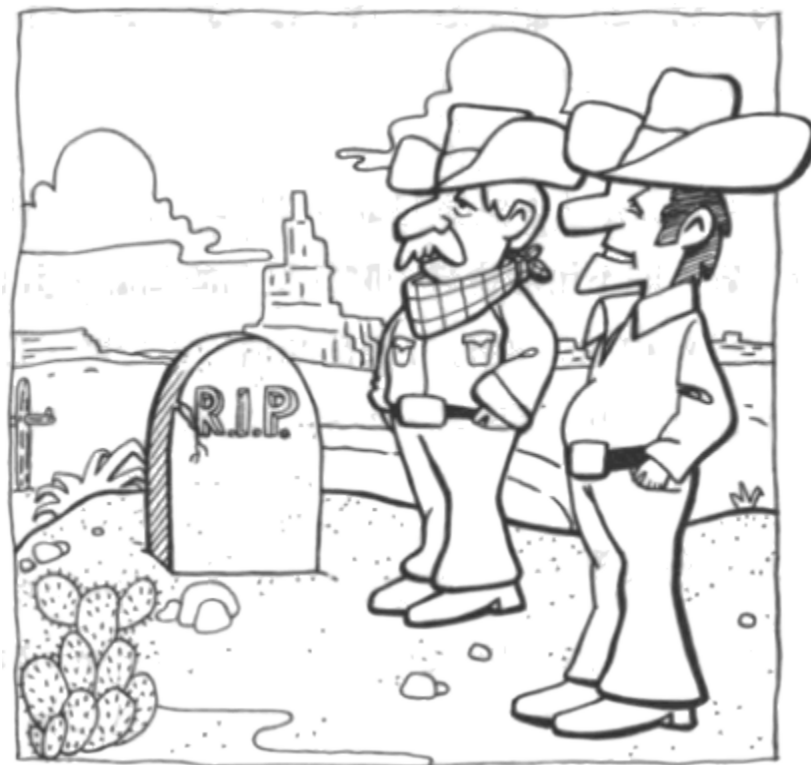
# 資料的離度：分散性

# 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



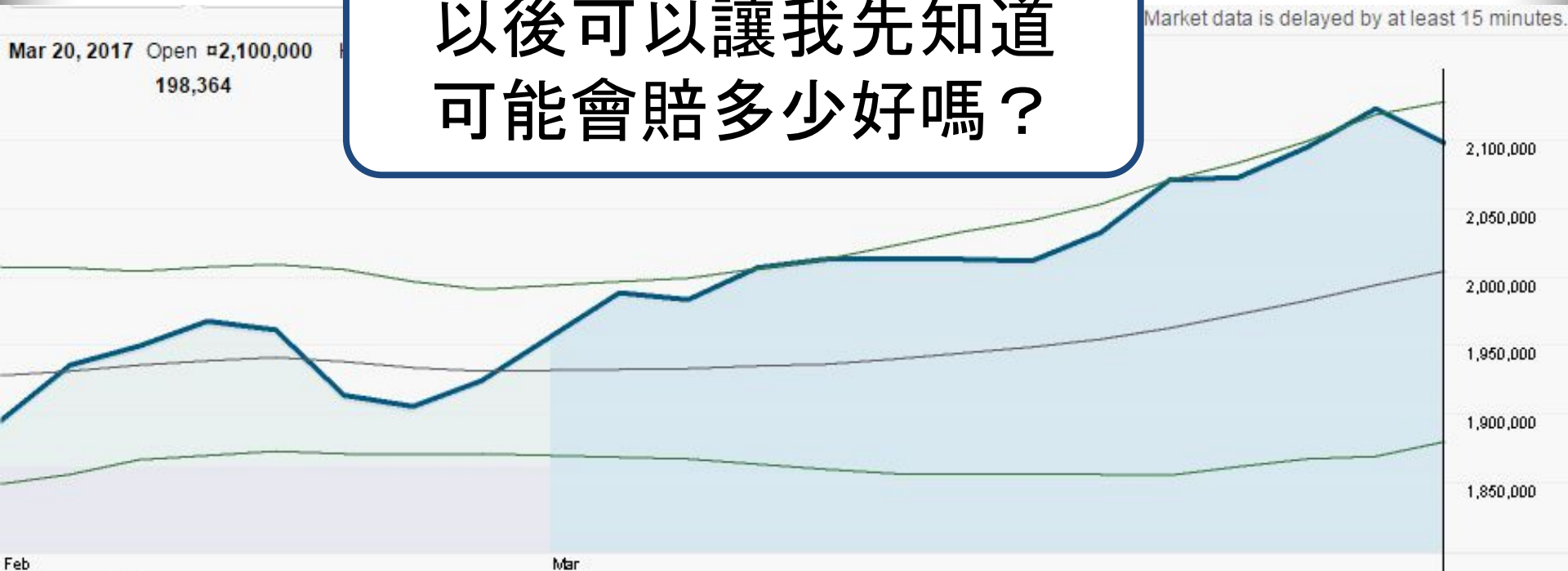
「我虧損了**26%**，怎麼會這樣？」

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。

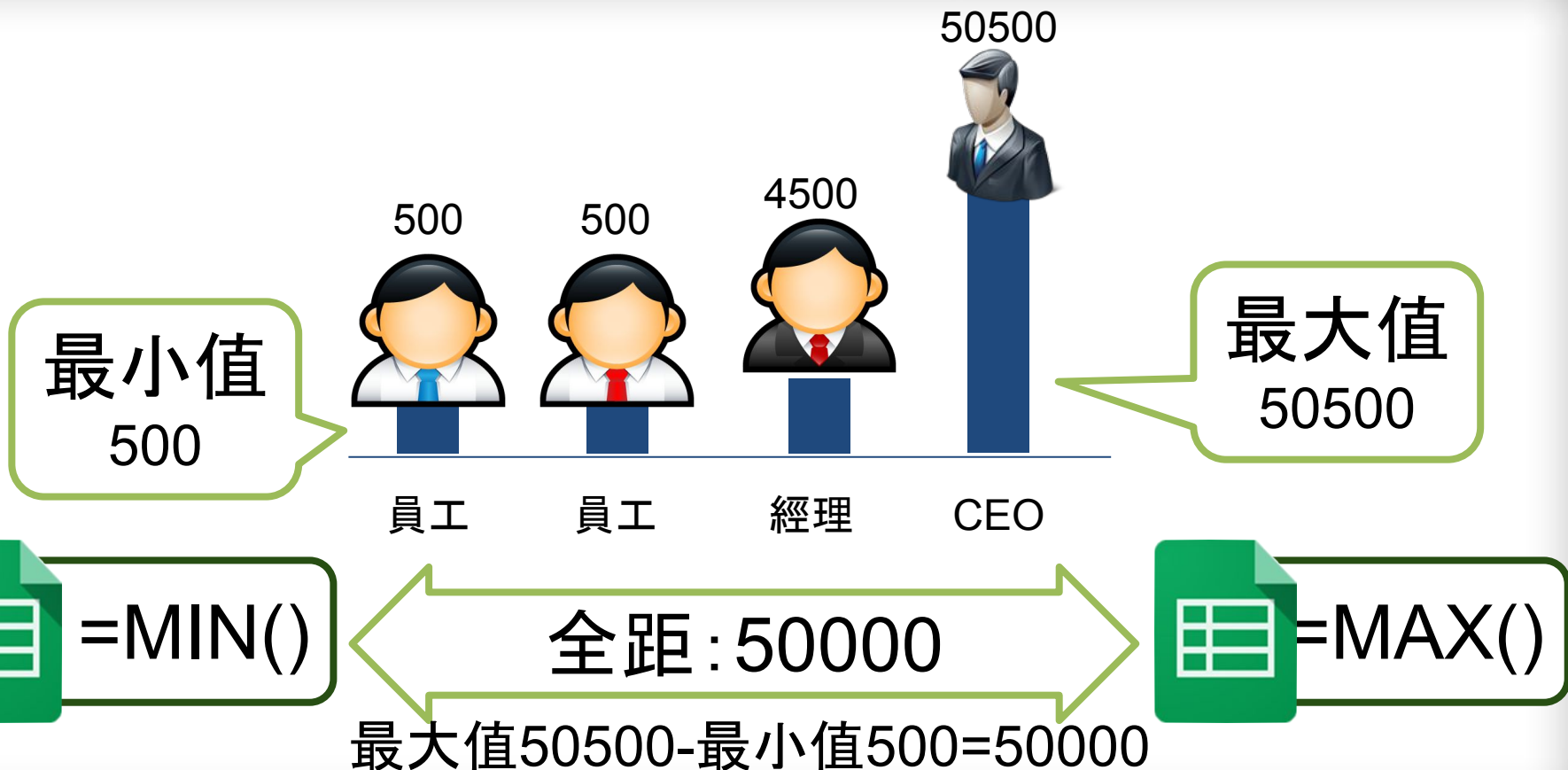




以後可以讓我先知道  
可能會賠多少好嗎？



# 全距:最大值-最小值



# 全距的問題

離群值

全距數：

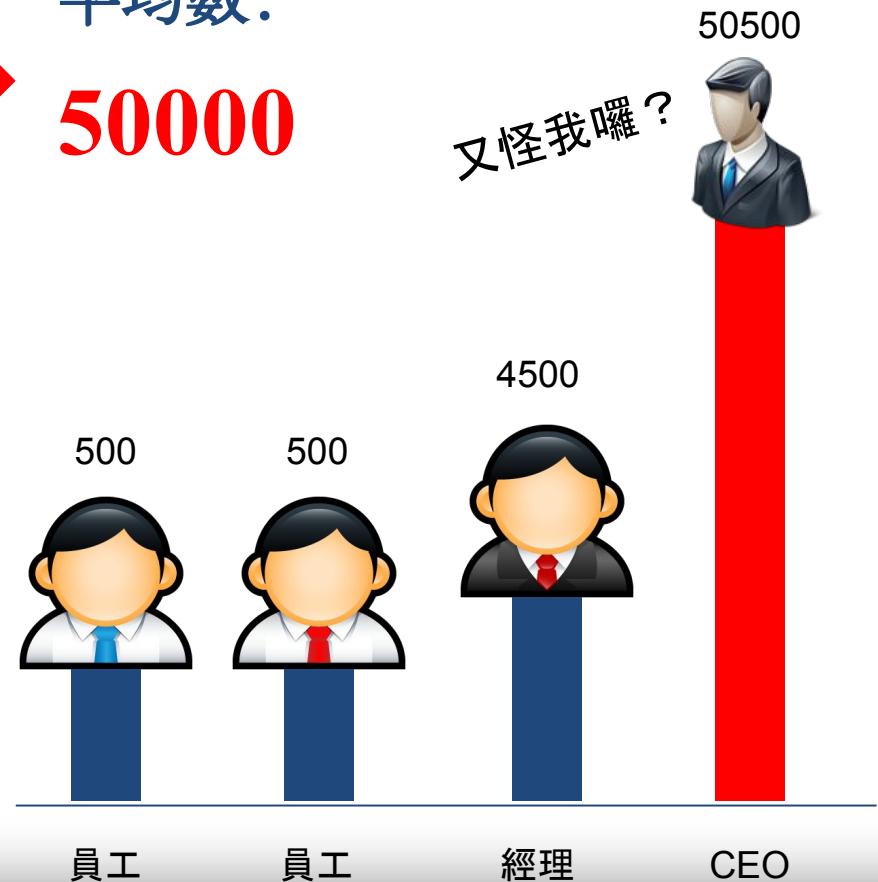
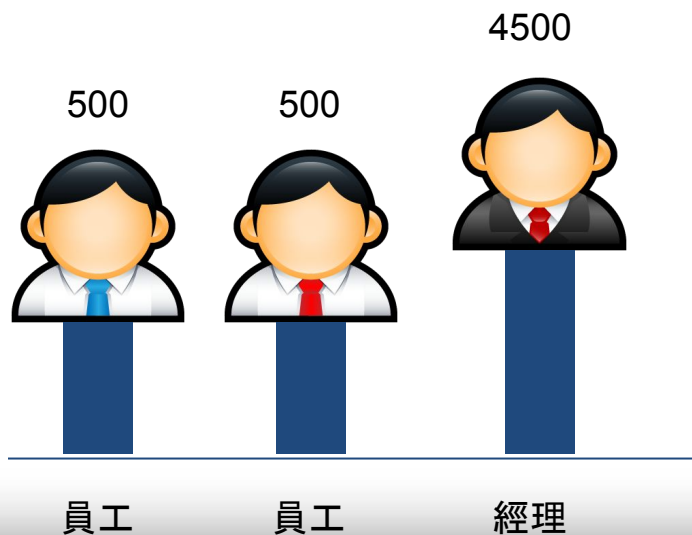
4000

平均數：

50000

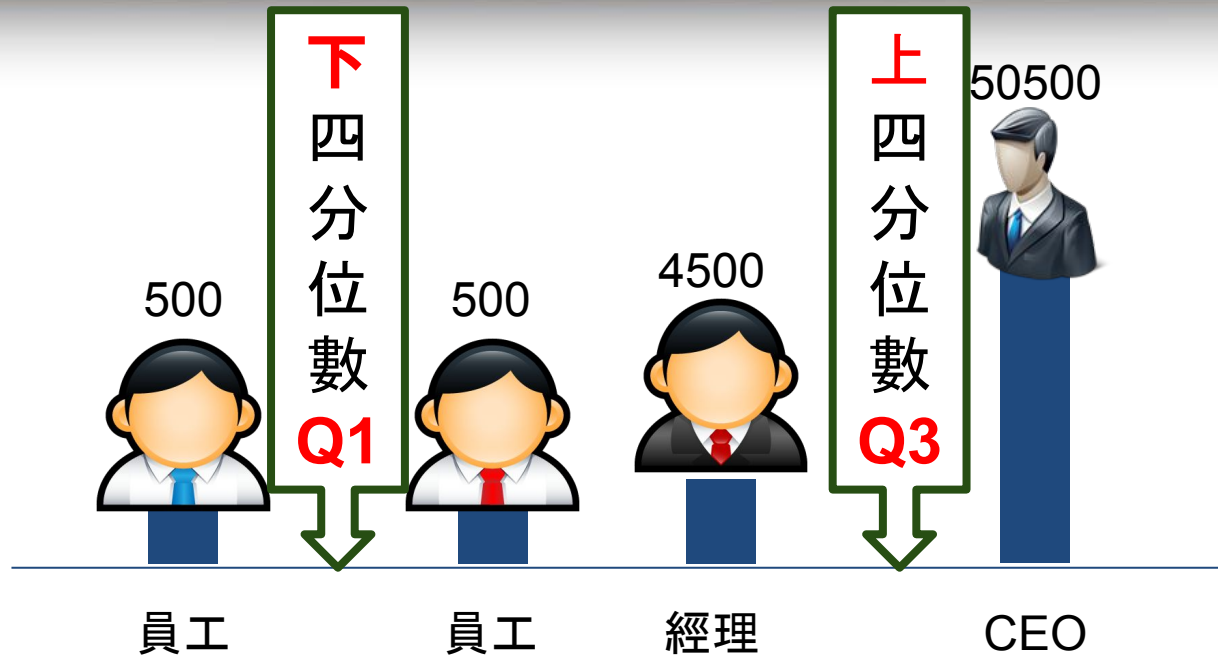
50500

又怪我囉？





# 內四分位距：上四分位數-下四分位數

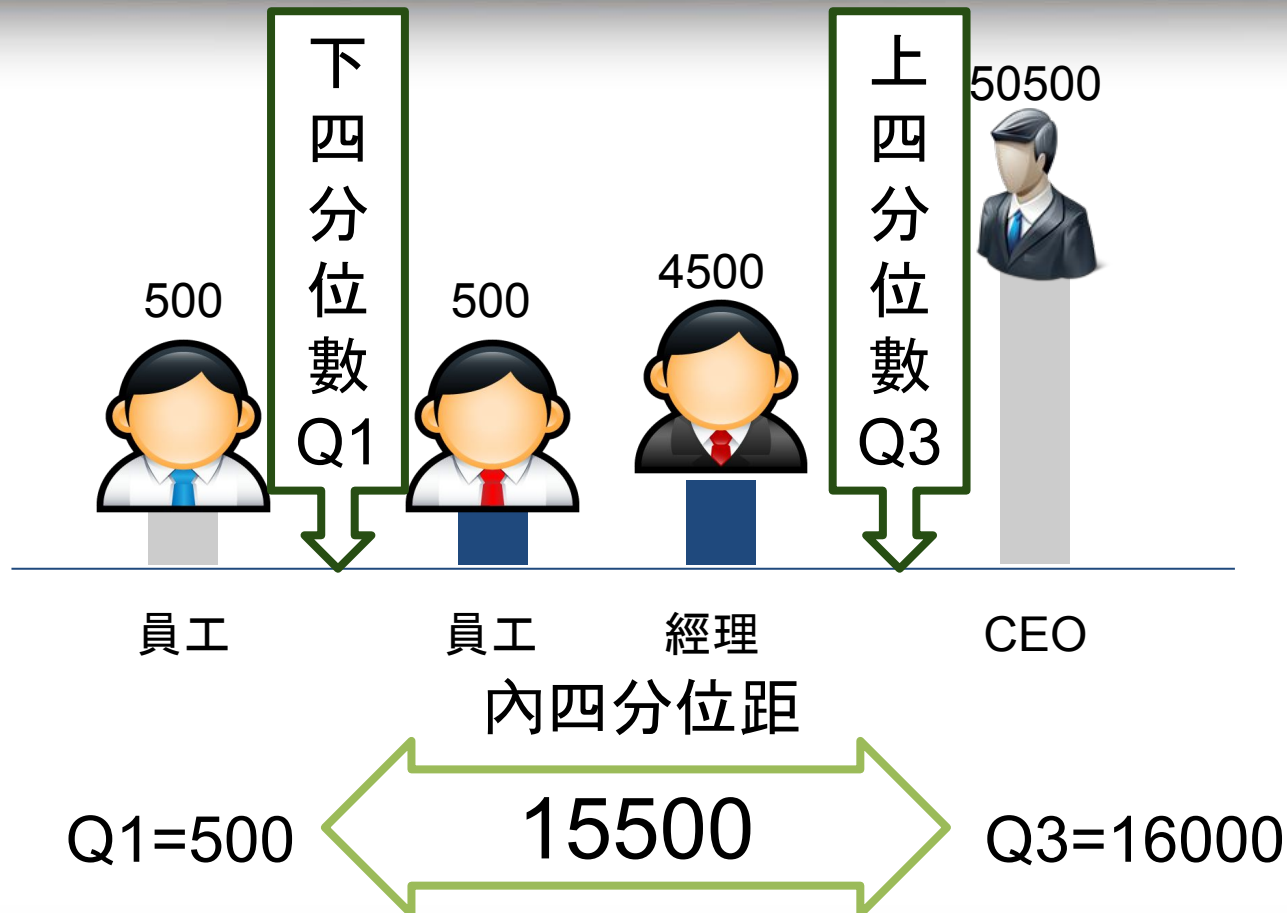


=QUARTILE(資料範圍,1)



=QUARTILE(資料範圍,3)

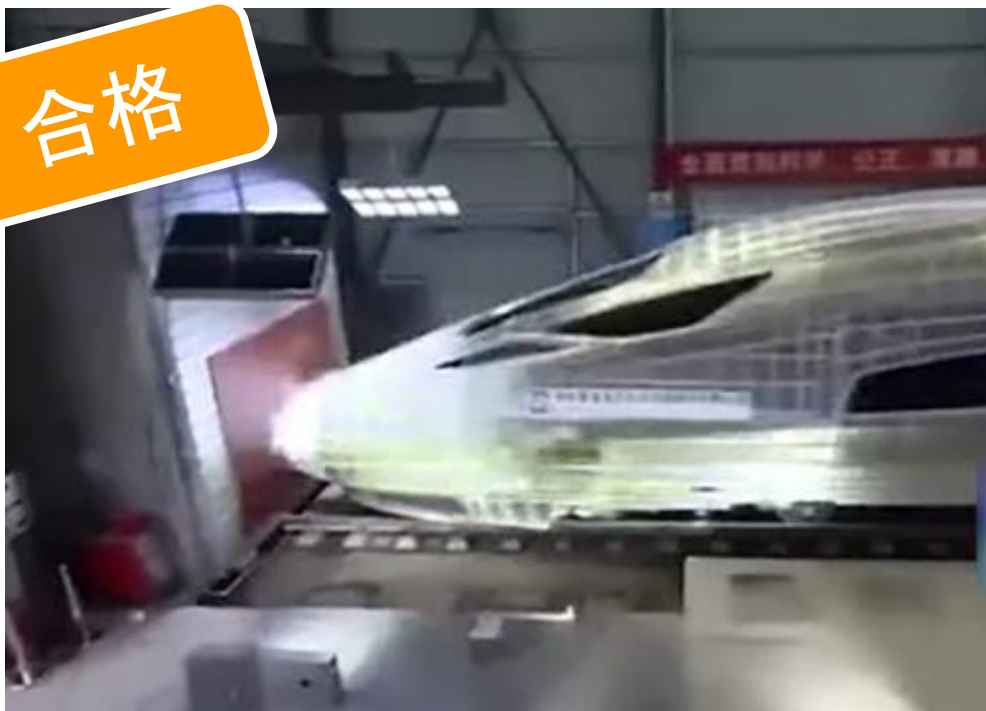
# 內四分位距：上四分位數-下四分位數



# 中國動車對撞測試

測試時速：**38公里**

合格



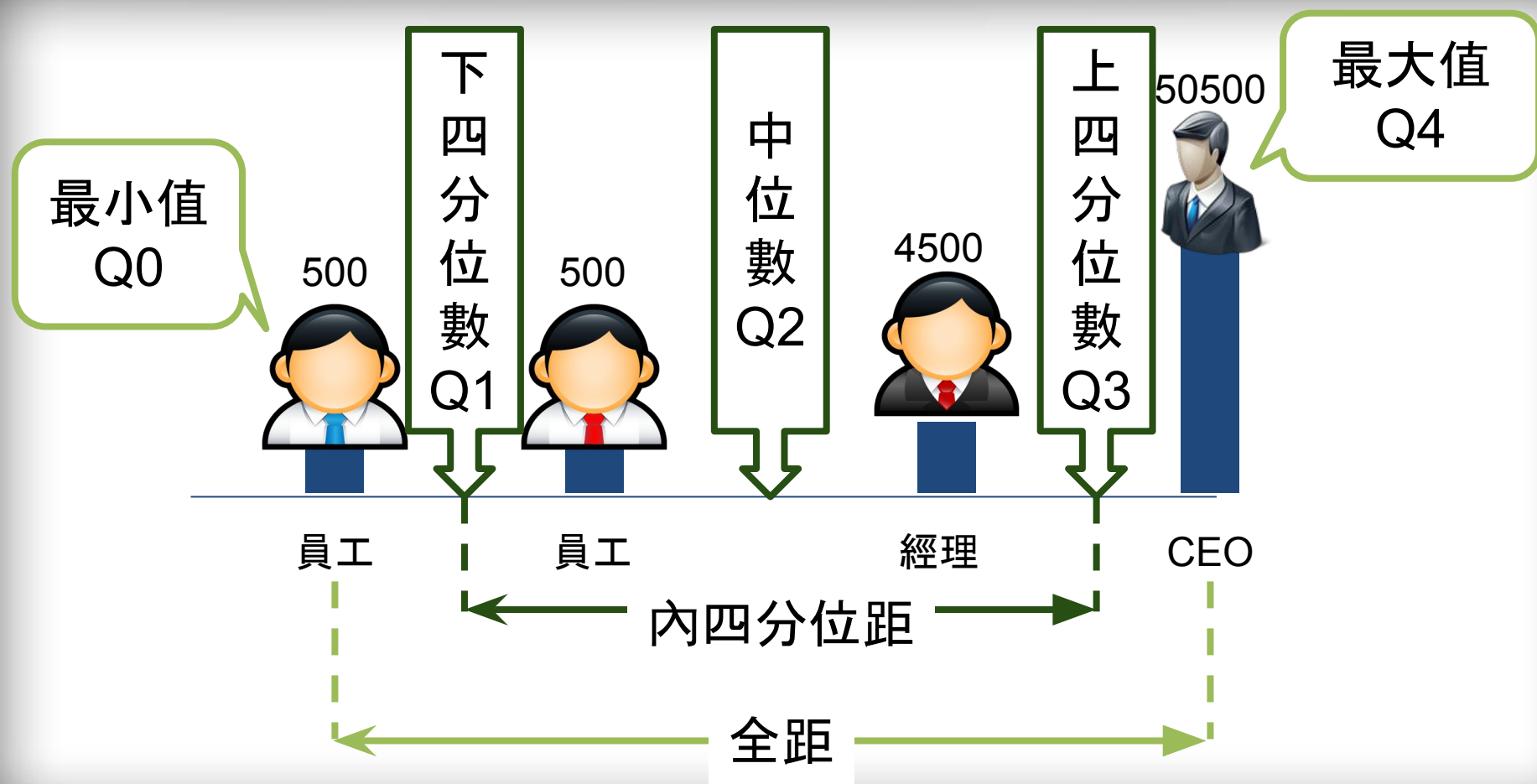
溫州動車對撞40死  
相對時速**83公里**  
只是特例，沒事的



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

更多選項

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	14		

data

# 汽車的價格

汽車案例數量:100筆



# W05-b. 五數綜合的計算

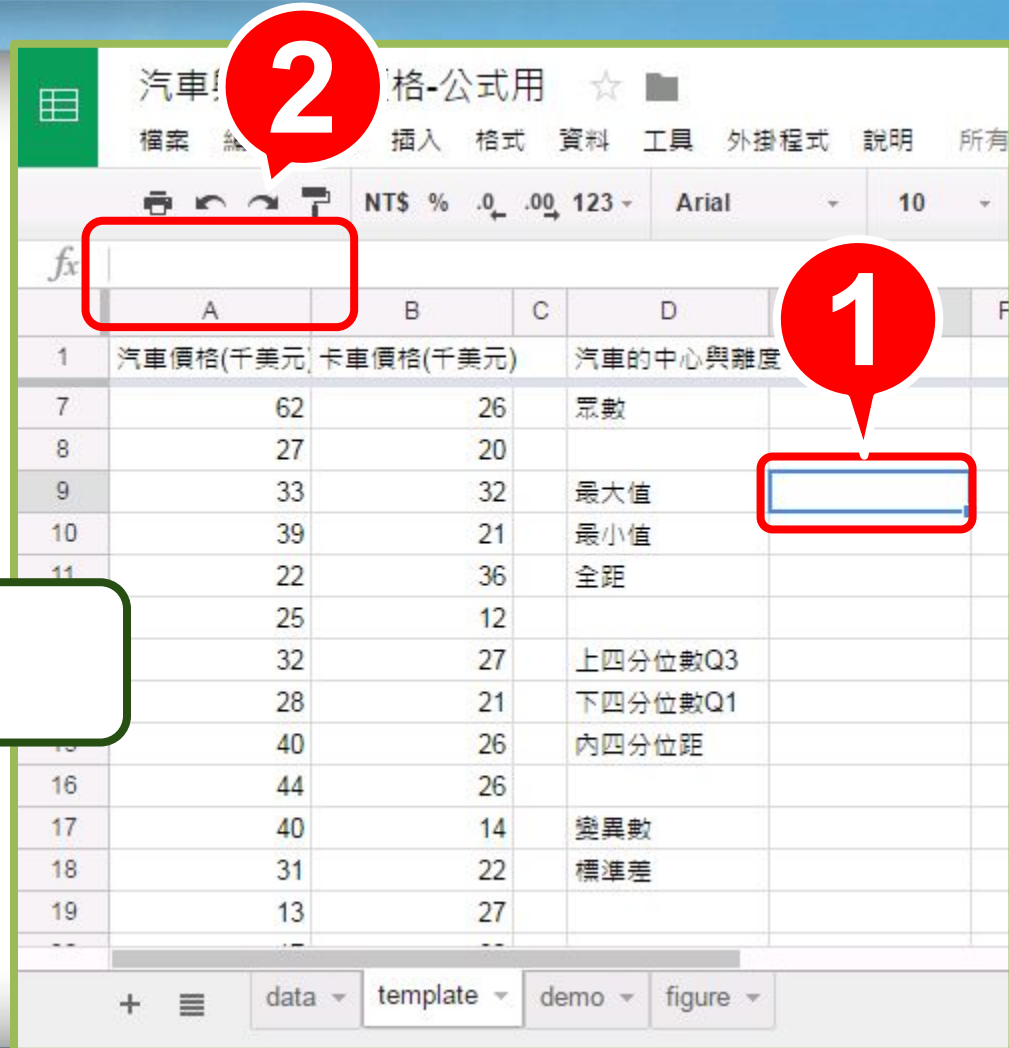
1. 插入最大值的公式與資料範圍
2. 插入最小值的公式與資料範圍
3. 計算全距
4. 插入上四分位數Q3的公式與資料範圍
5. 插入下四分位數Q1的公式與資料範圍
6. 計算內四分位距



# 1. 插入最大值的公式與資料範圍

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

**=MIN(A2:A101)**

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
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
	25	12		
	32	27		上四分位數Q3
	28	21		下四分位數Q1
	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		

Red circle 1 points to the empty cell next to '最小值' in column D. Red circle 2 points to the empty formula bar at the top.

### 3. 計算全距

1. 點選「全距」  
旁邊的空白格 (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)

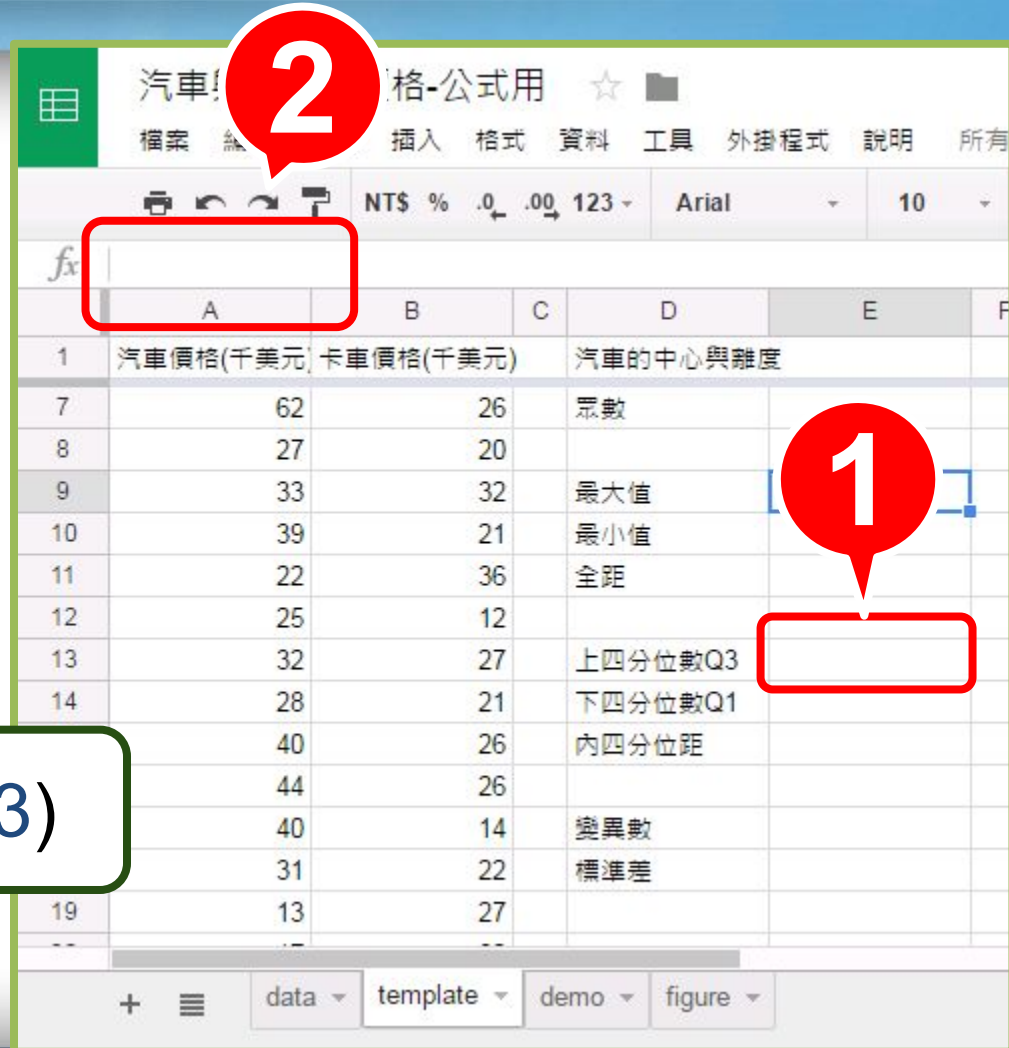
Excel spreadsheet showing data for car prices and statistics. The formula bar (labeled with a red '2') is highlighted, and the cell E11 (labeled with a red '1') is highlighted in the spreadsheet.

	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」  
旁邊的空白格 (E13)
2. 點選「公式列」  
的空白處
3. 輸入公式:

**=QUARTILE(A2:A101,3)**



	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	
	40	26		內四分位距	
	44	26			
	40	14		變異數	
	31	22		標準差	
19	13	27			



## 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」  
旁邊的空白格 (E14)
2. 點選「公式列」  
的空白處
3. 輸入公式:

**=QUARTILE(A2:A101,1)**

The screenshot shows an Excel spreadsheet with the following data:

	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	
	40	26		內四分位距	
	44	26			
	40	14		變異數	
	31	22		標準差	
19	13	27			

## 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

$$=E13-E14$$

Excel spreadsheet showing data for car and truck prices, and statistical calculations. Red callouts 1 and 2 highlight the formula bar and cell E15 respectively.

	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-... ☆ pulipuli.chen@gmail.com

檔案 編輯 檢視 註解 共用 所有變更 更多選項

fx  $=E13-E14$

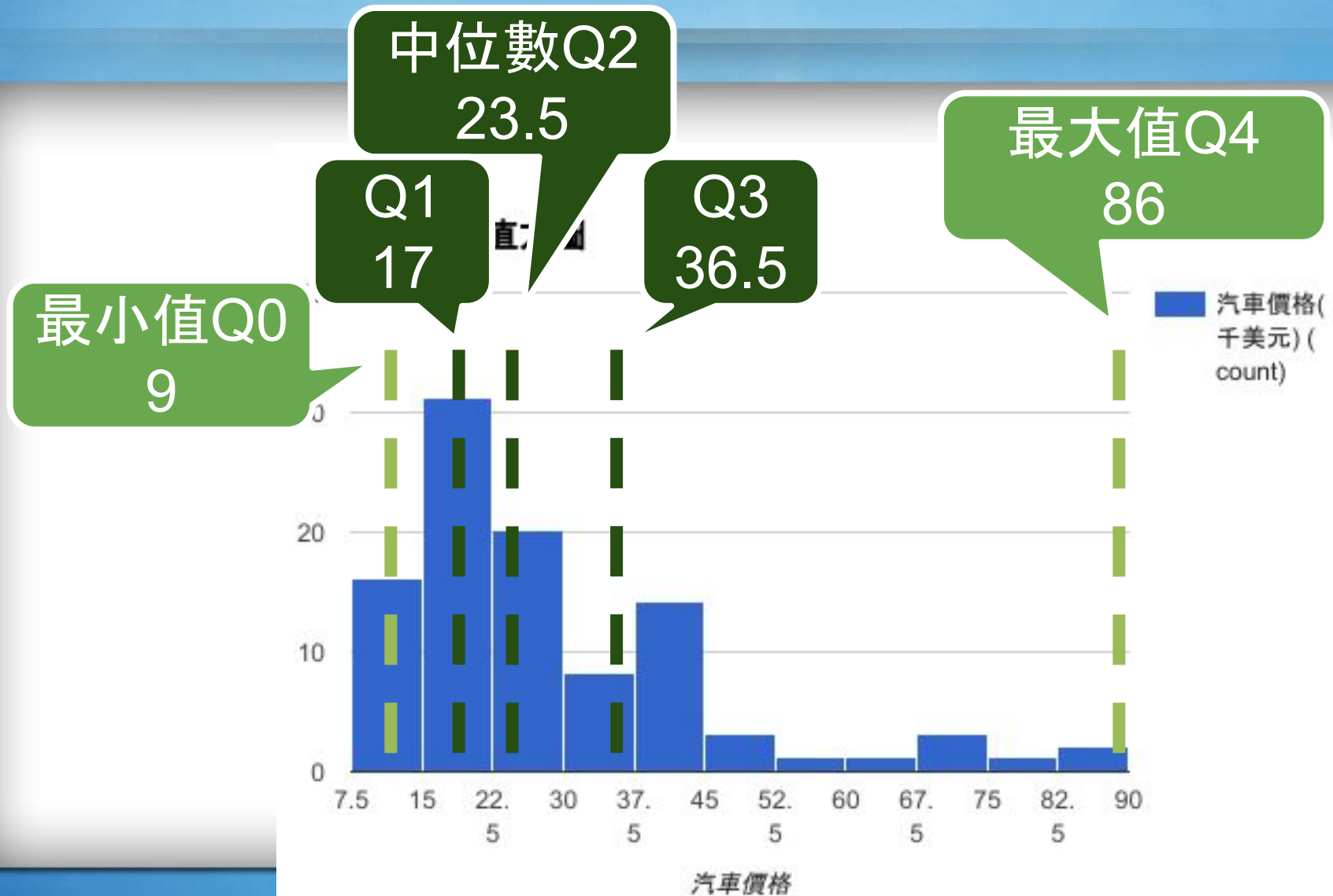
	A	B	C	D	E	F
1	汽車價格(千美元)	卡車價格(千美元)	汽車的中心與離度			
7	62	26	眾數		22	
8	27	20				
9	33	32	最大值		86	
10	39	21	最小值		9	
		36	全距		77	
		12				
		21	上四分位數Q3		36.5	
			下四分位數Q1		17	
15	40	26	內四分位距		19.5	
16	44	26				

+ ≡ data ◀ ▶

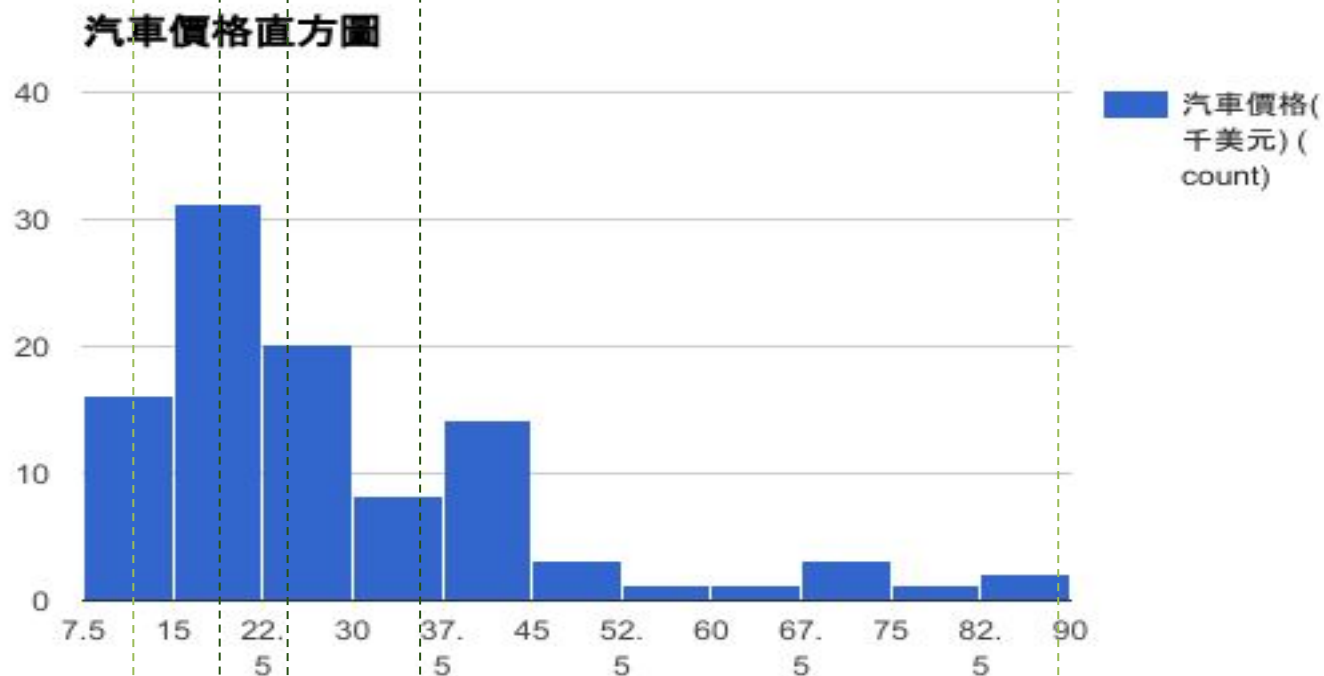
Part 5.

## 實作：五數綜合的繪圖

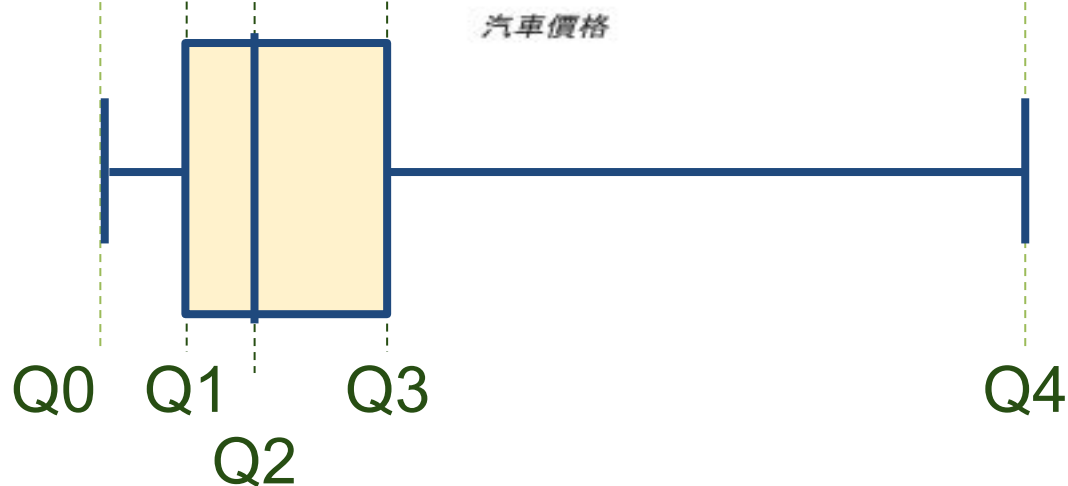
# 觀察五數綜合在資料分佈的位置



直方圖

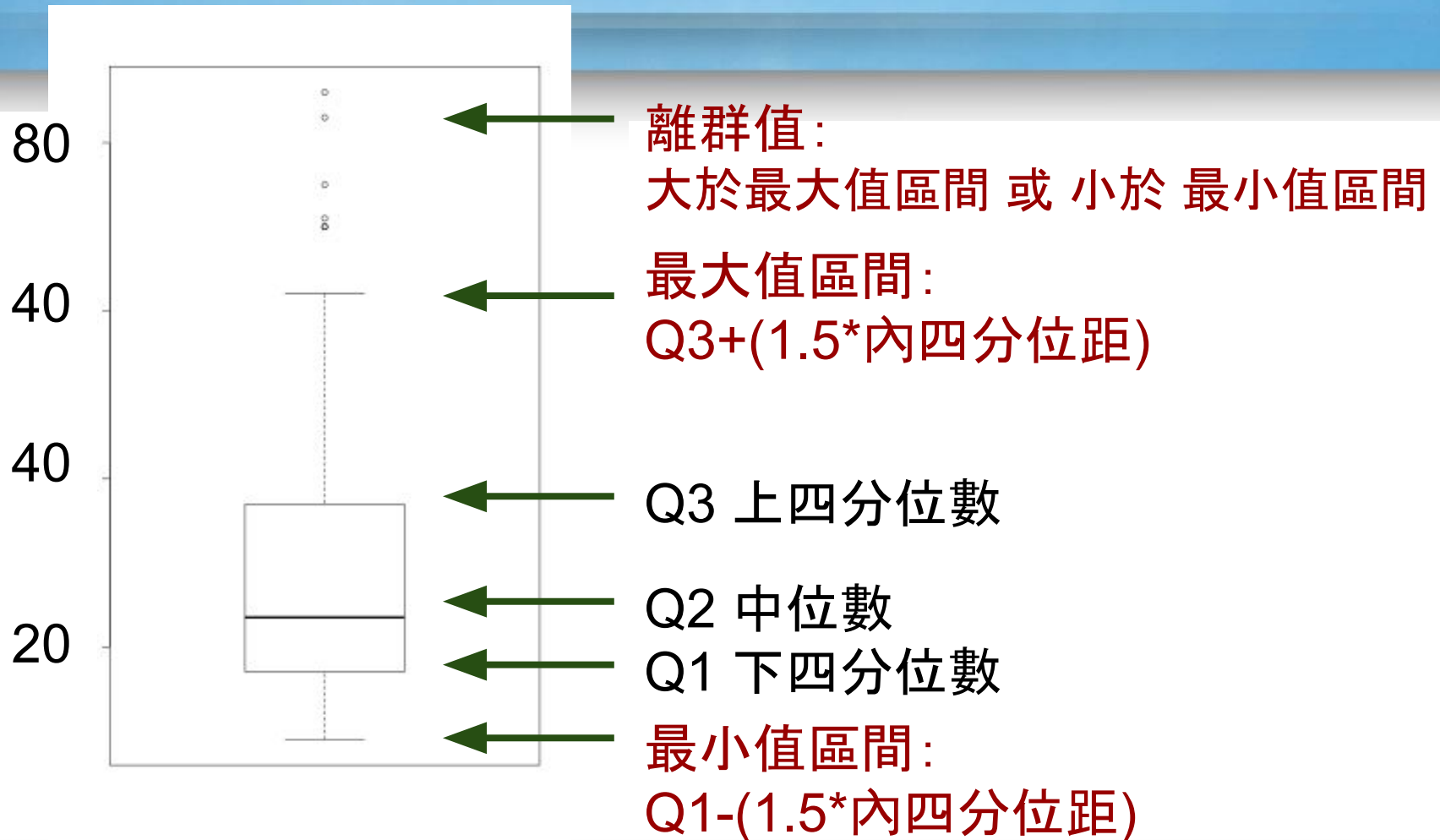


箱型圖  
(盒鬚圖)





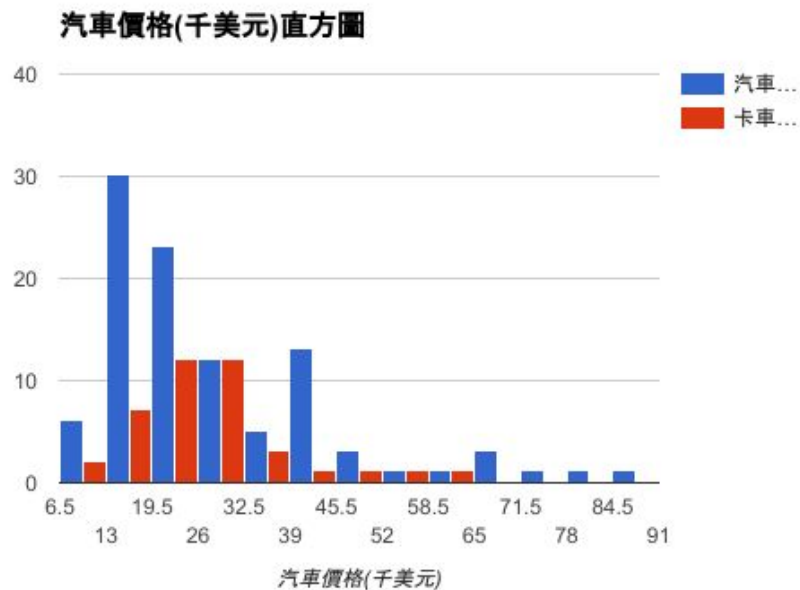
# R的箱型圖 (盒鬚圖)



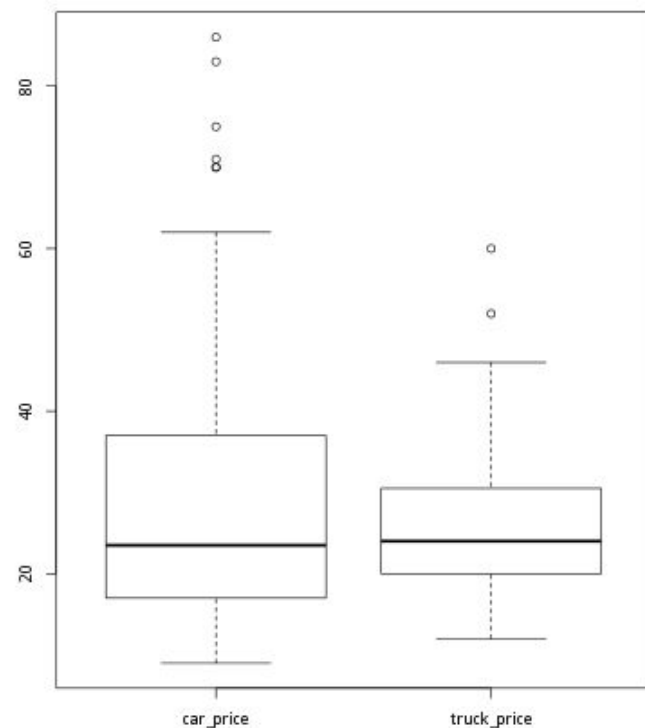
# 箱型圖的特色

## 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

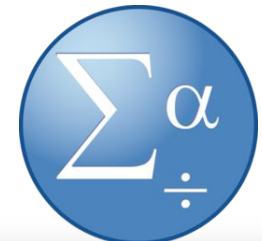
## R-Web Base

- 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

檔案 編輯 檢視 插入 格式 資料

NT\$ % .0\_ .00\_ 123

	A	B	C
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	
13	32	27	
14	28	21	
15	40	26	

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





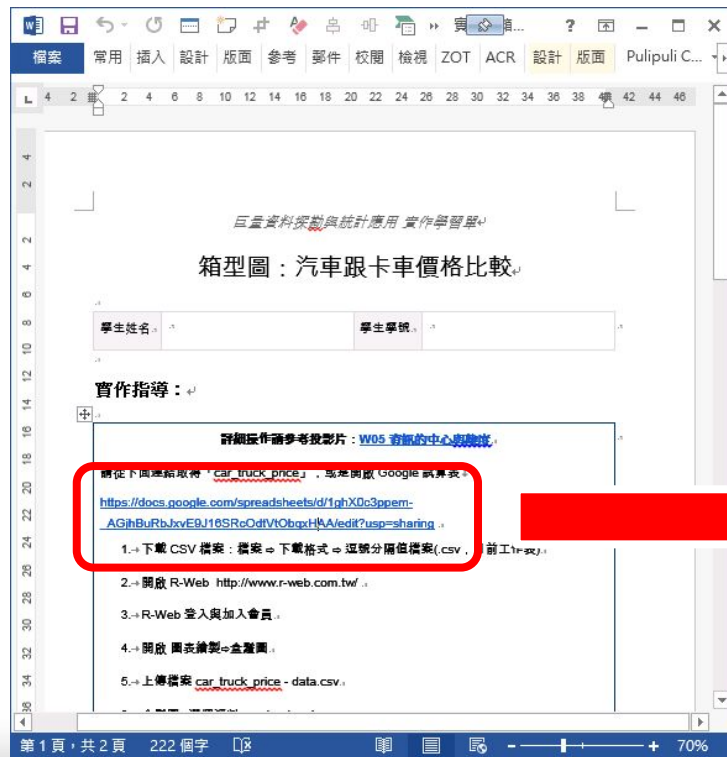
# W05-c. 箱型圖： 汽車跟卡車價格比較

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

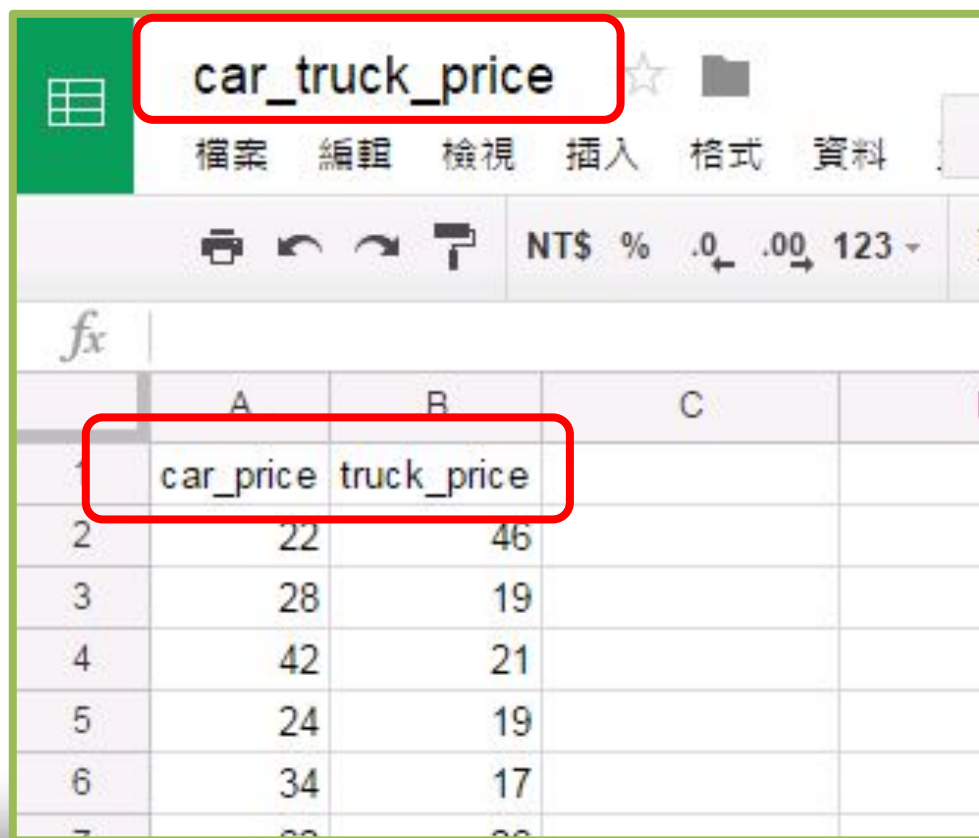


car\_truck\_price

	A	B	C
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

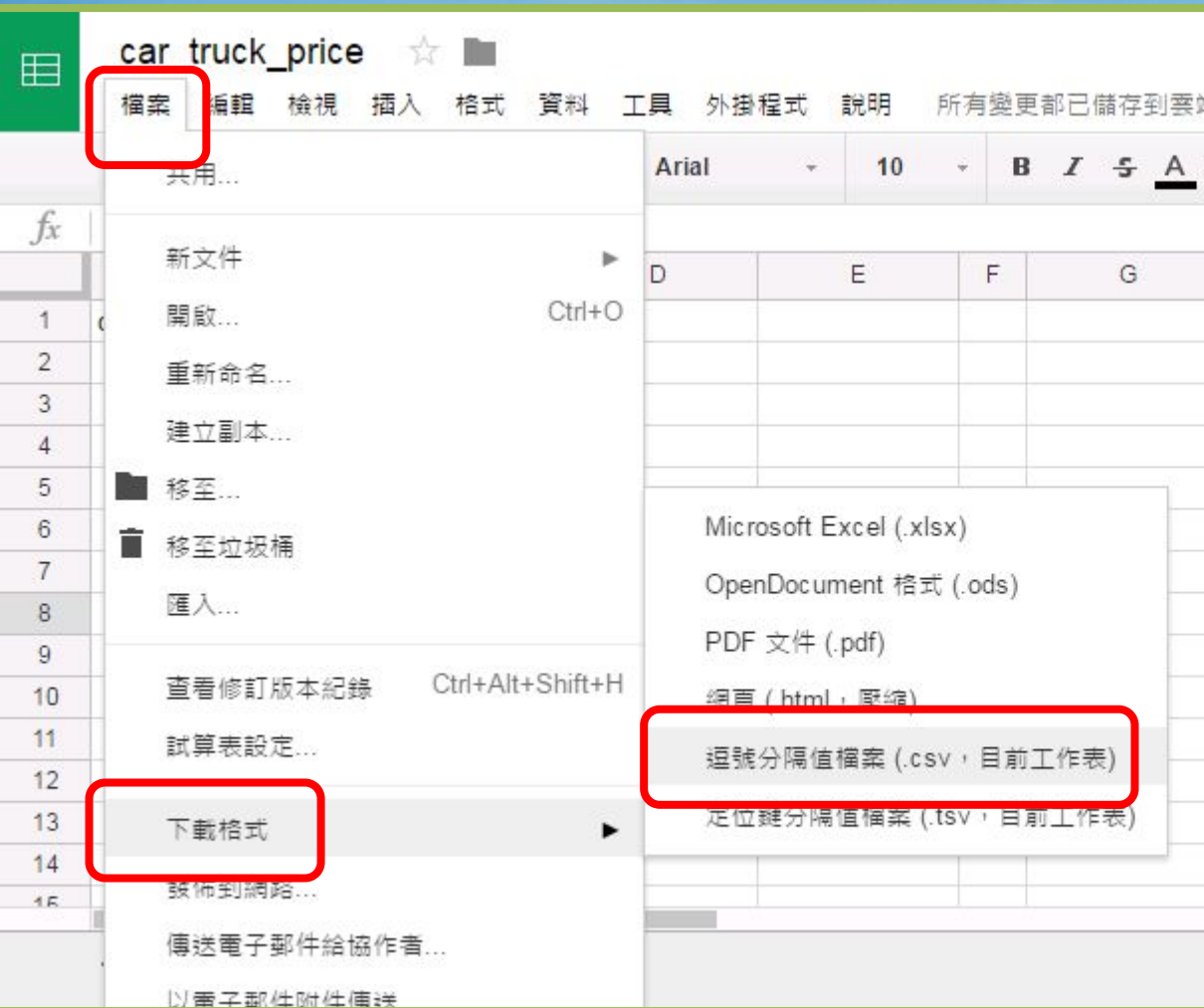
英文、數字、\_ (底線)、- (橫線)



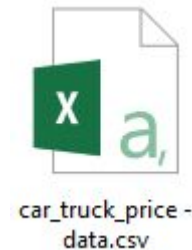
The screenshot shows an Excel spreadsheet with a table containing two columns of variable names: 'car\_price' and 'truck\_price'. The formula bar at the top displays 'car\_truck\_price', which is a concatenation of the two column headers. Both the table header row and the formula bar text are highlighted with red rectangles to illustrate valid naming conventions.

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

# 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)



## 2. 開啟R-Web

R-Web

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

The screenshot shows the homepage of the R-Web Data Analysis & Guiding System (DAGS-C). The header includes the system name, a navigation menu, and user options. The main content area features a '使用流程簡介' (Usage Process Introduction) section with a flowchart and a 'DAGS-C' logo. The flowchart illustrates the process from data input to output, involving steps like '資料未處理' (Data Not Processed), '資料處理功能' (Data Processing Function), '系統導引功能' (System Guidance Function), '分析方法功能' (Analysis Method Function), and '輸出結果功能' (Output Result Function). The footer contains a '最新消息' (Latest News) section and a '雲端計算聯盟' (Cloud Computing Alliance) link.

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 | 初階導引

初階導引 | 資料處理 | 分析方法 | 圖表編製 | 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**

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



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 ▾

f 談 分享

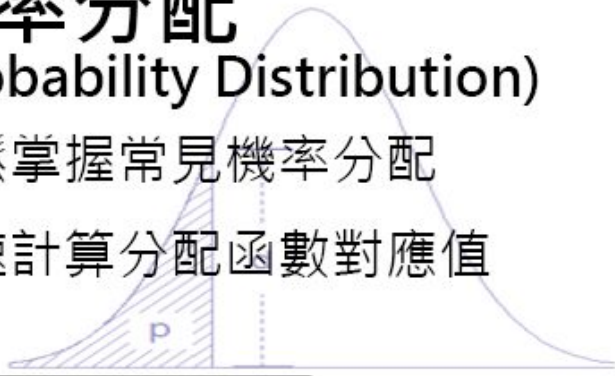
G+

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

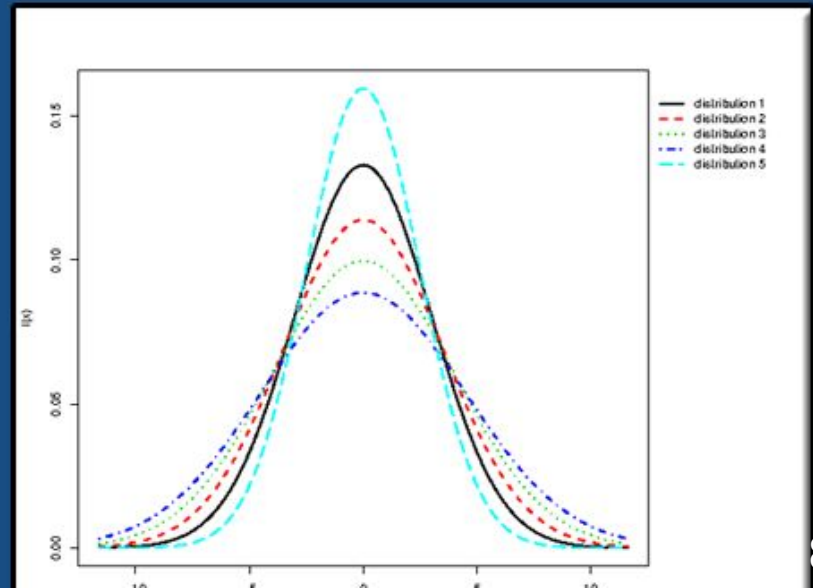
## 機率分配 (Probability Distribution)

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

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

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 自創巨集

首頁 » 加入會員

加入會員  
Join RDAGS Member

會員資料設定:

\* 所有欄位皆為必填資料

一、會員帳號資料

會員帳號 Member ID	<input type="text"/>	4~12字元，且不含特殊字元
會員密碼 Password	<input type="password"/>	8~12字元，且不含特殊字元
確認密碼 Confirm Password	<input type="password"/>	請再次輸入密碼

二、個人基本資料

電子信箱 E-mail Address	<input type="text"/>		系統將寄發驗證信至此信箱，一組電子信箱帳號僅可申請一個會員帳號
西元出生年 Year of Birth	<input type="text"/>	請選擇	
性別	<input type="radio"/>	<input type="radio"/>	

## 4. 開啟 圖表繪製⇨盒鬚圖

首頁 ⇨ 圖表繪製 ⇨ 盒鬚圖

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 | 初階導引

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

次數分配表  
列聯表  
莖葉圖  
散佈圖  
曲線(面)圖  
直方圖  
長條圖  
圓餅圖  
盒鬚圖

Graph for 20 rules  
資料採礦(Data Mining)  
關聯規則分析-Apriori method  
從雜亂無章的數據中去蕪存  
挖掘出具有高度相關性的資

Scatter plot for 345 rules

www.r-web.com.tw/chart/step1.php?method=box\_chart

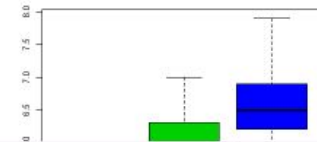
# 5. 上傳檔案 (1/5)

## 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(開新視窗)

== 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 範例說明 影音教學

無論是欲分析的資料或是輔助分析的額外資料，均可透過此功能上傳至系統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：

選擇檔案 car\_truck\_price - data.csv

car\_truck\_price - data.csv

確認上傳 檢視所有已上傳之檔案

**確認上傳**

雲 版權所有 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
- 資料中的遺失值表示符號為何：



5.	34	17
6.	62	26

回上一步 下一步驟

**下一步驟**

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 (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

car\_truck\_price

變數名稱

car\_price

truck\_p

變數型態

數值

數值

1.

22

46

2.

28

19

3.

42

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

\* 使用者若對資料有隱私、安全性上之考量，建議可將變數名稱以代碼取代。

回上一步

確認儲存

確認儲存

# 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 談 分享

G+

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

								依	資料檔名稱 ▾	遞增 ▾	排序
	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"/>	DGC				10	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(開新視窗)

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)

PCS3

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)

car\_price  
truck\_price

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)

首頁 » 輸出結果 » 統計圖表：盒鬚圖

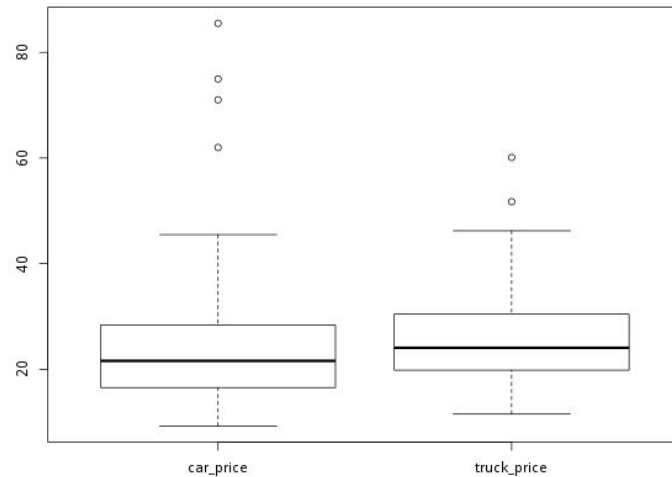
輸出結果

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：



# W05-c. 箱型圖： 汽車跟卡車價格比較

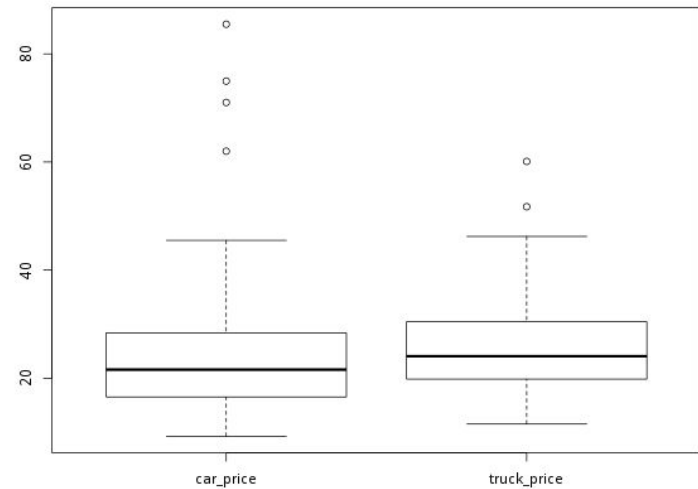
上機實作

參數

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

• 盒型圖：

箱型圖



Part 6.

# 資料的離度：變動性



# 心臟沒這麼大顆

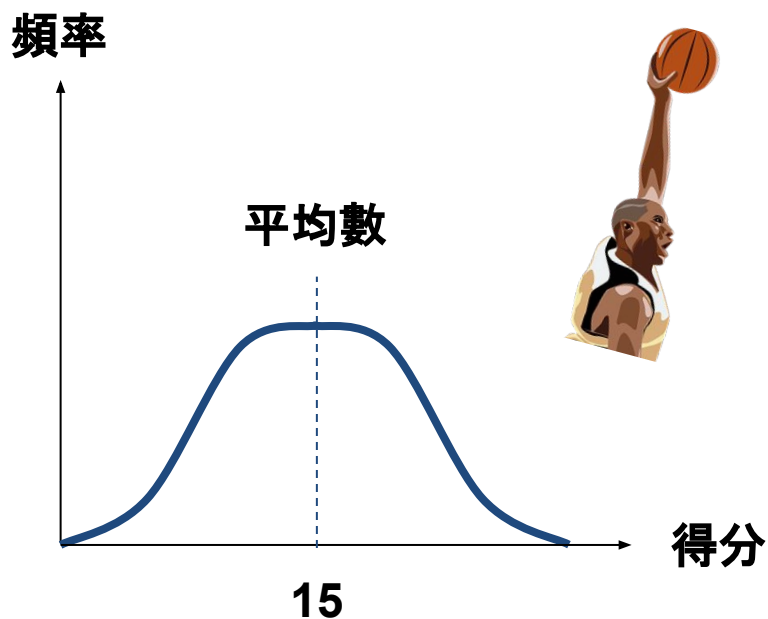


「我想要穩定賺錢就好，  
不想要高風險，可以嗎？」

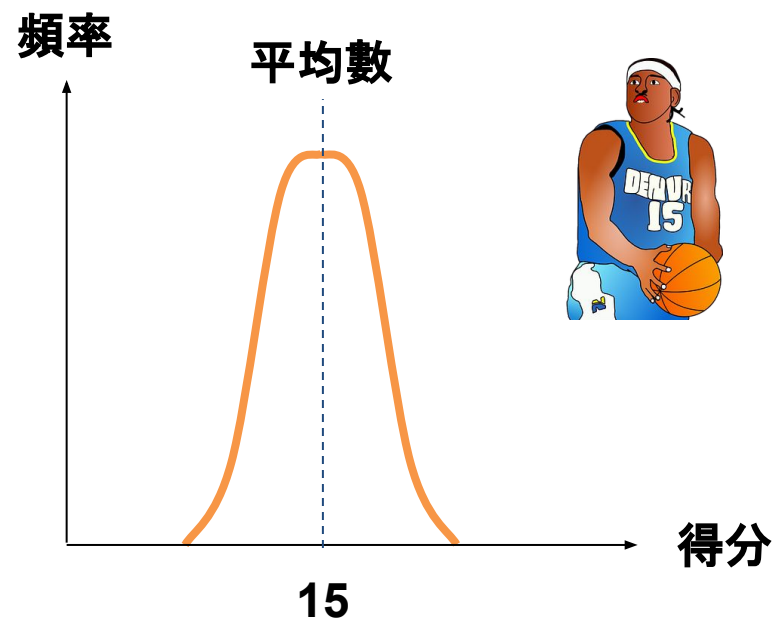


# 教練也想知道該派誰好？

球員阿布的得分

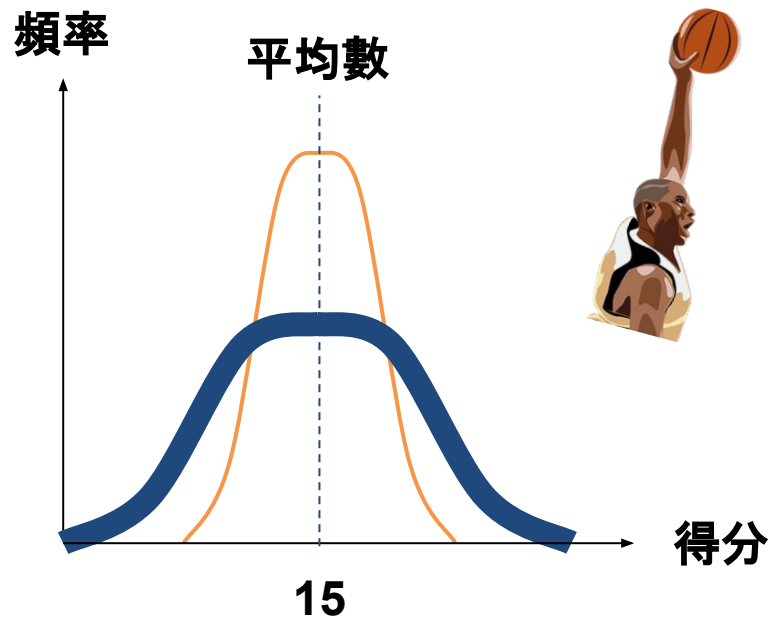


球員阿丁的得分



「那個球員比較能預測他下一場的表演呢？」

# 球員阿布的得分

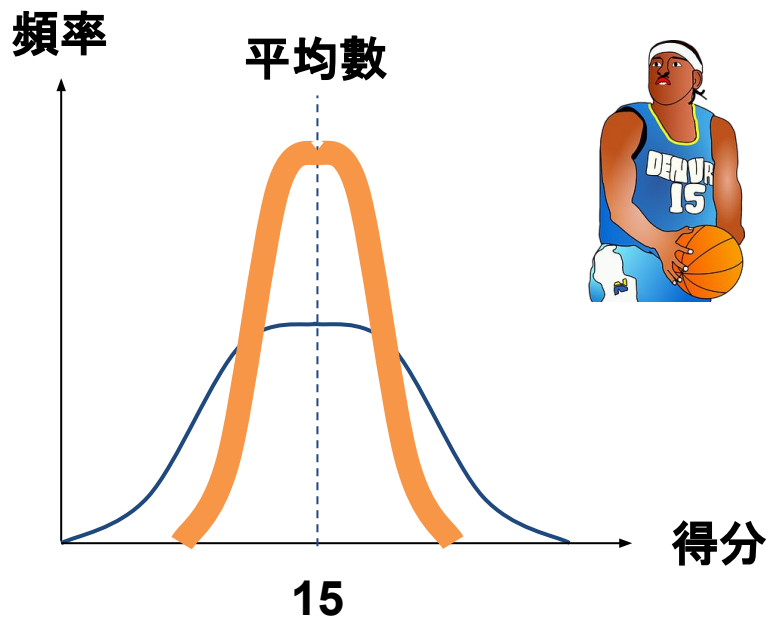


- 資料比較分散
- 離平均數比較遠
- 狀況好時會得高分
- 狀況不好時可能很糟

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)

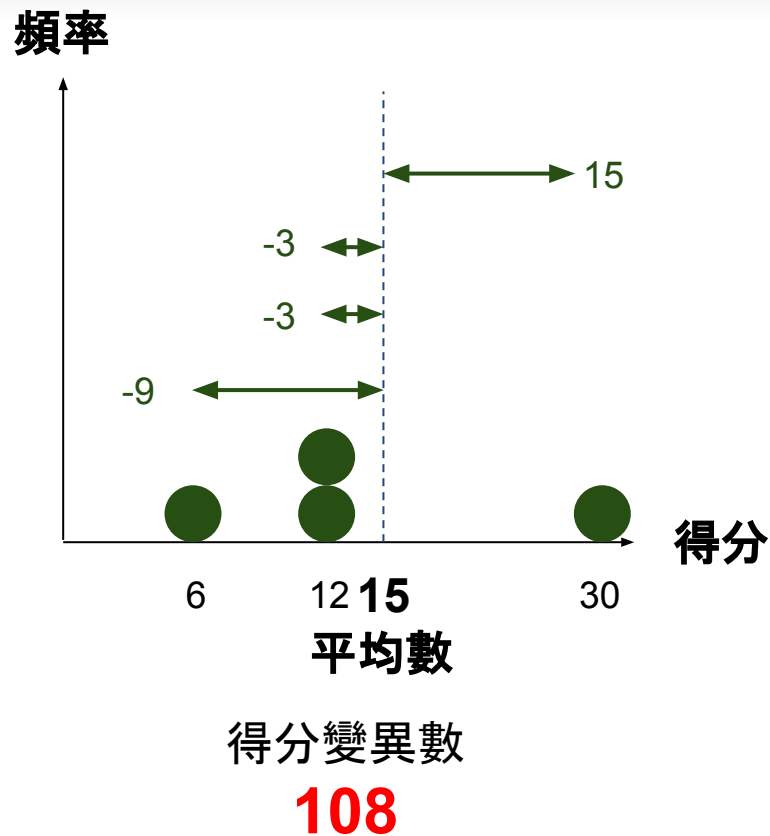


**資料  
變動性**

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



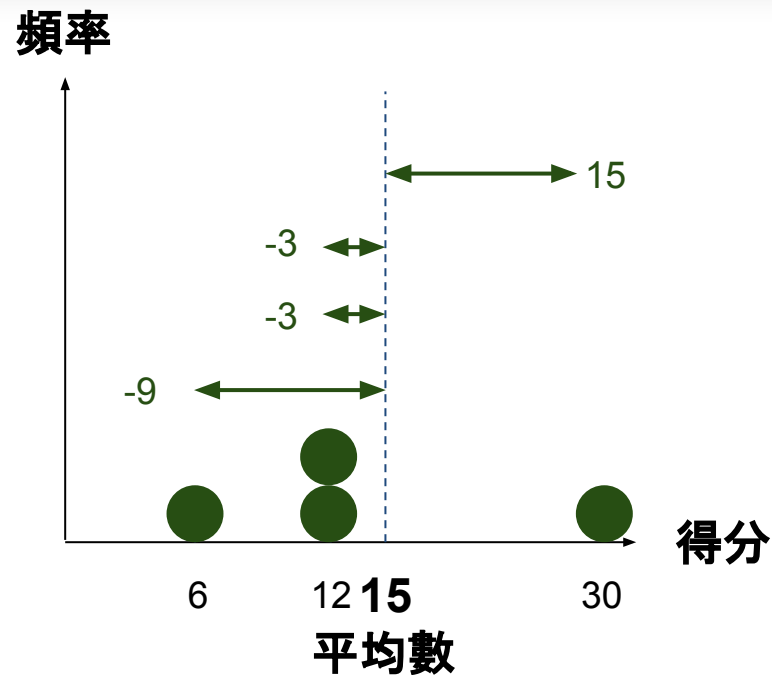
變異數是呈現所有數值與平均數的平均概念

- 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)



**=VAR()**

# 變異數的限制



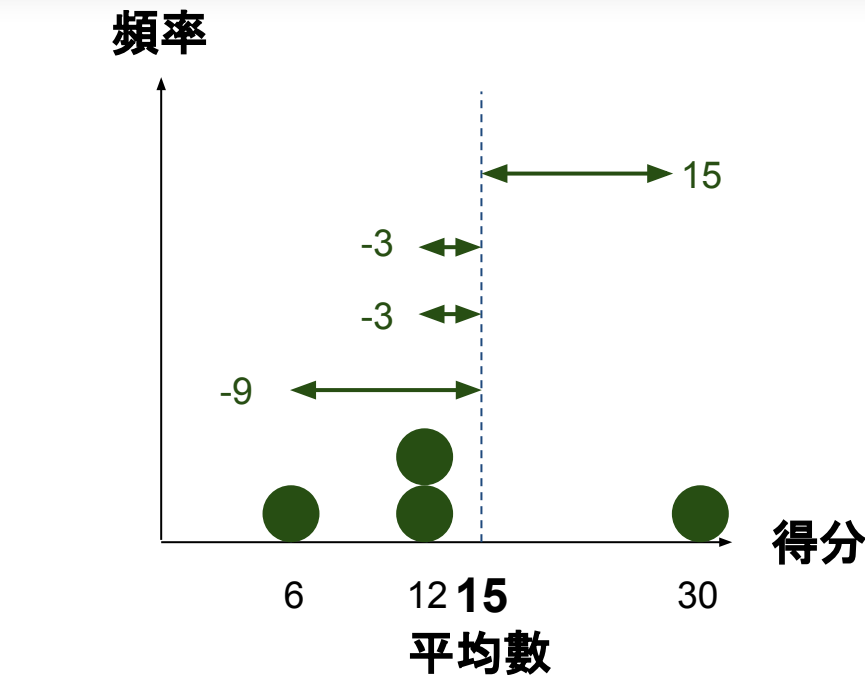
得分變異數

108

哪裡像距離平均？



# 資料變動性的第二招 標準差



得分變異數  
**108**

取平方根 $\sqrt{\quad}$

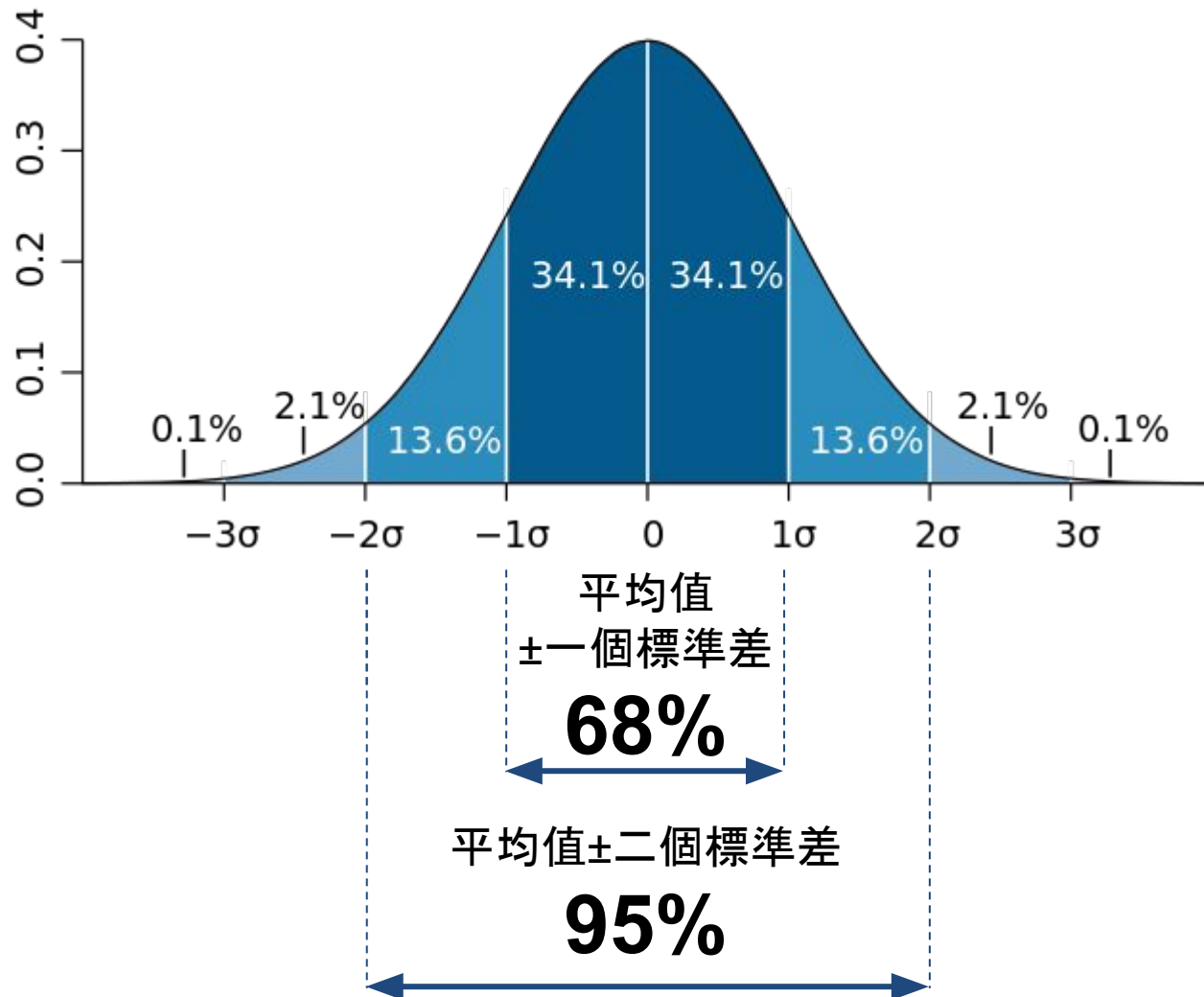
得分標準差

**10.39**

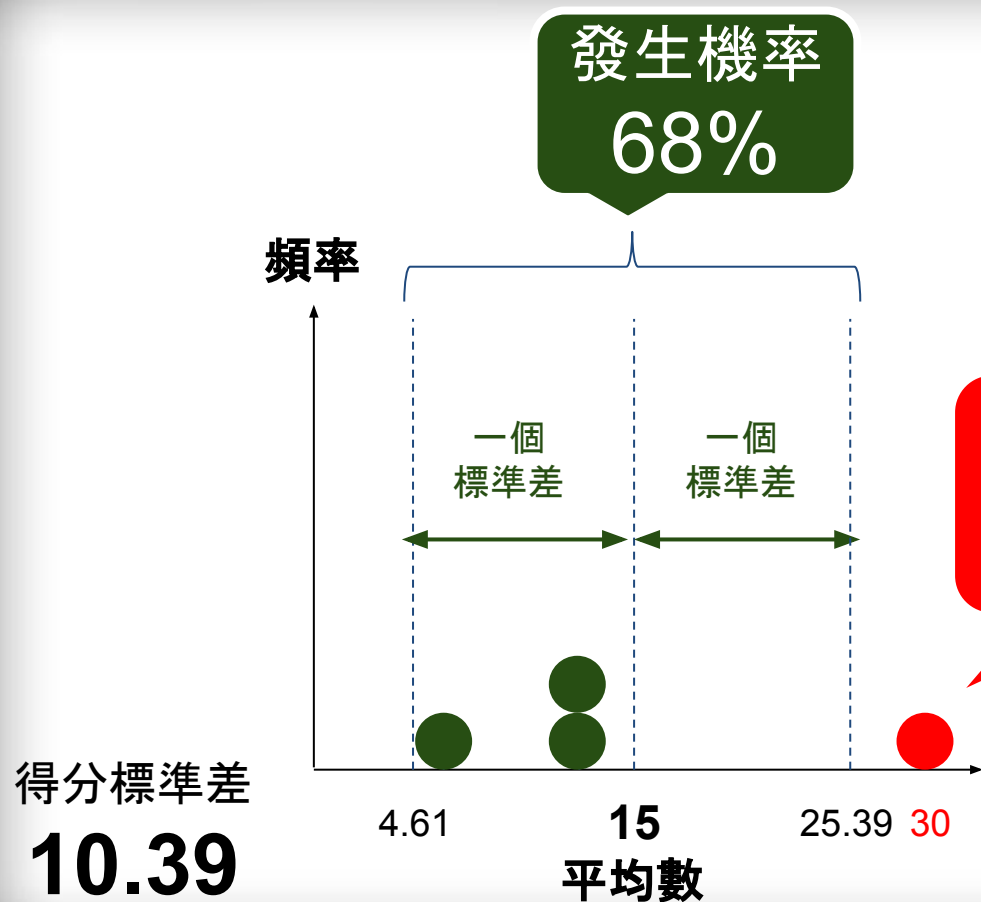
很像  
距離平均

# 常態分配的規則

## 用標準差預測發生機率



# 得分30的機率？

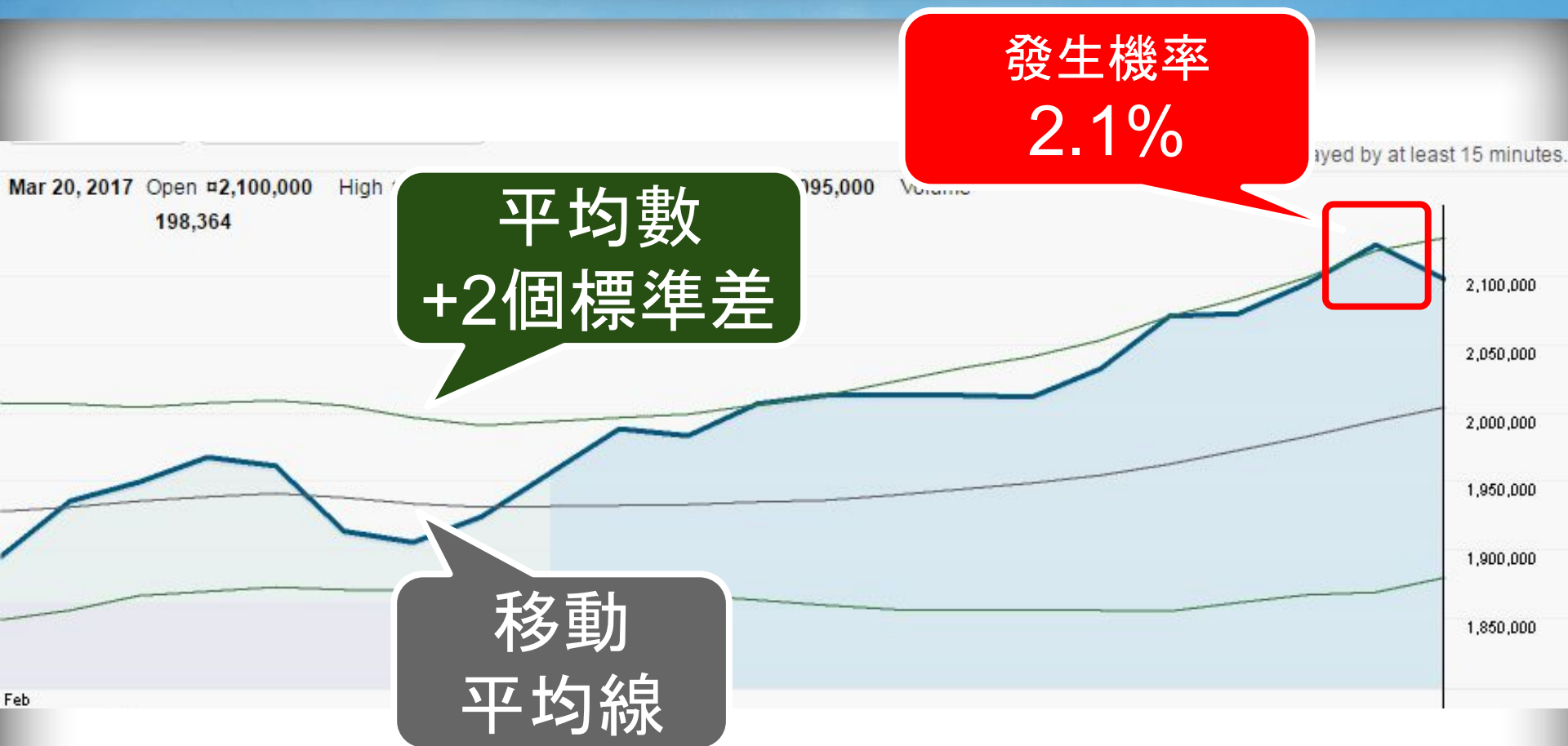


教練！  
我這次肯定能拿高分！

發生機率  
13.6%



# 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

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



=STDEVP()

樣本標準差

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

貝索校正

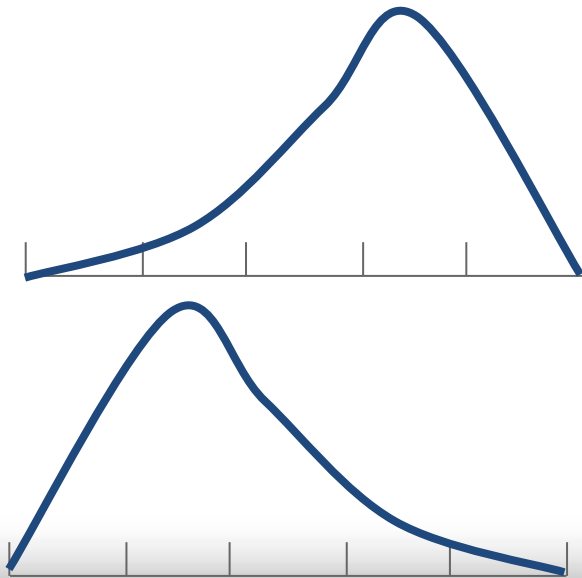


=STDEV()

# 如何選擇適當的綜合數 值描述？

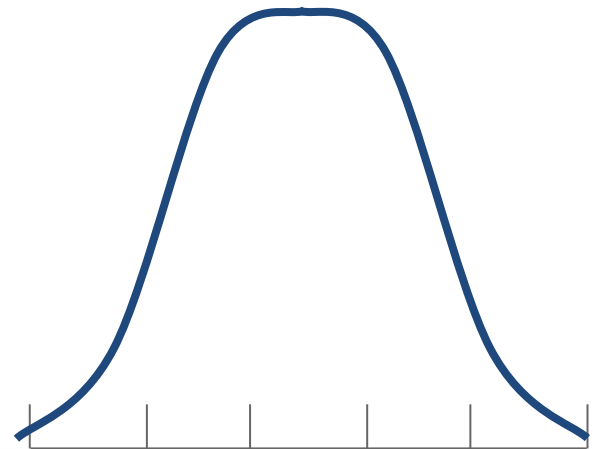
- 資料有偏斜分佈、長尾巴
- 資料有離群值

選擇「五數綜合」



- 資料分佈對稱(常態分配)
- 沒有離群值

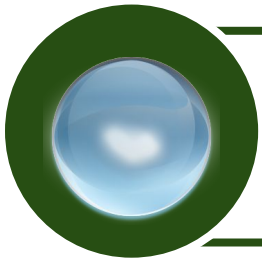
選擇「平均數 & 標準差」



# 資料敘述級的步驟

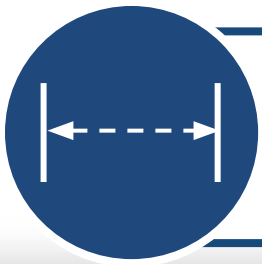


**統計圖表 (全貌概況)**



**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... 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

更多選項

fx 汽車價格(千美元)

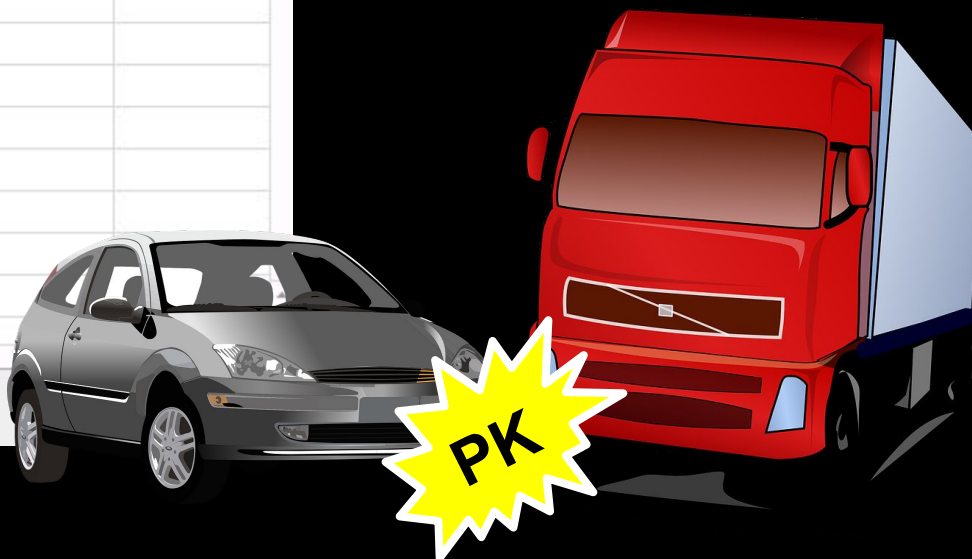
	A	B	C	D
1	汽車價格(千美元)	卡車價格(千美元)		
2	22	46		
3	28	19		
4	42	21		
5	24	19		
6	34	17		
7	62	26		
8	27	20		
9	33	32		
10	39	21		
11	22	36		
12	25	12		
13	32	27		
14	28	21		
15	40	26		
16	44	26		
17	40	14		

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



# W05-d.

## 變異數與標準差的計算

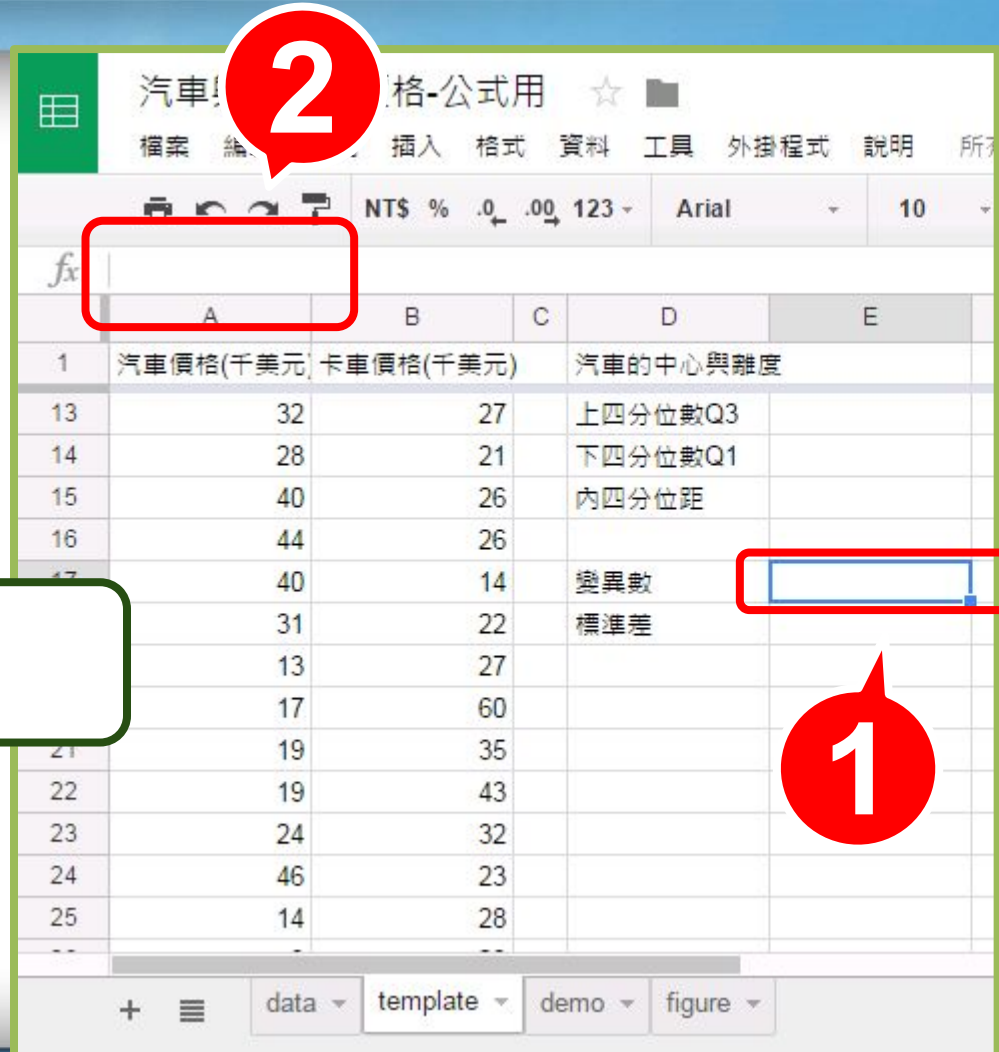
1. 插入汽車變異數的公式與資料範圍
2. 插入汽車標準差的公式與資料範圍
3. 插入卡車的變異數與標準差的公式與資料範圍



# 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. 插入標準差的公式與資料範圍

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

**=STDEV(A2:A101)**

The screenshot shows an Excel spreadsheet with the following data:

	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			

The formula bar is highlighted with a red circle and the number 2. Cell E18 is highlighted with a red circle and the number 1.



### 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

**=VAR(B2:B41)**

- 卡車標準差

**=STDEV(B2:B41)**

汽車與卡車的價格-公式用

檔案 編輯 檢視 插入 格式 資料 工具 外掛程式 說明 所有變更都已儲存到雲端硬碟

NT\$ % 0.00 123 Arial 10 B I A

	B	C	D	E	F	G	H
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
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							

data template demo figure

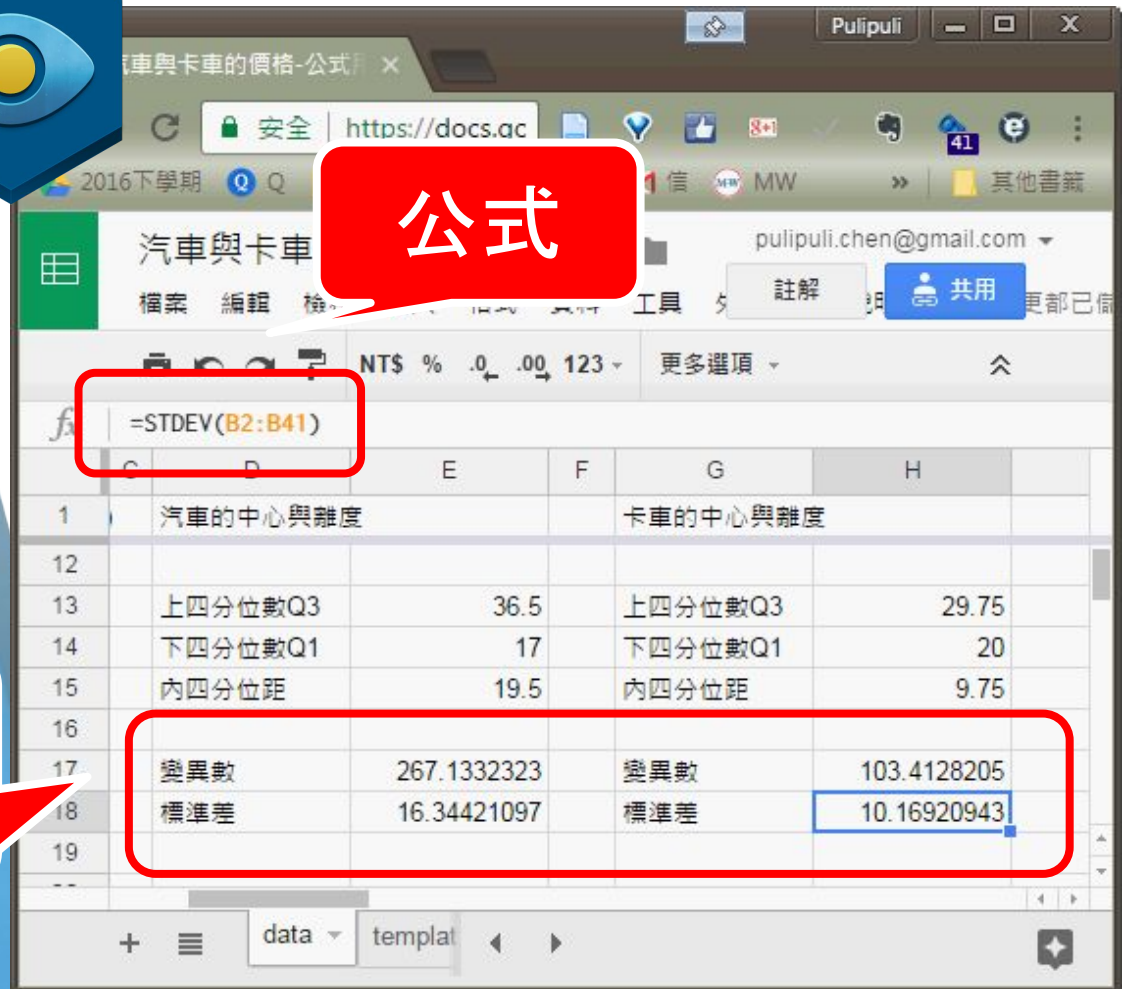
# W05-d.

## 變異數與標準差的計算

上機實作

公式

變異數  
標準差



	汽車	卡車
1	汽車的中心與離度	卡車的中心與離度
12		
13	上四分位數Q3	上四分位數Q3
14	下四分位數Q1	下四分位數Q1
15	內四分位距	內四分位距
16		
17	變異數	變異數
18	標準差	標準差
19		

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

檔案 編輯 檢視 插入 格式 資料 工具 外接程式 說明 所有窗 註解 共用

NT\$ % .0 .00 123 Arial 10 更多選項

$\text{fx}$  =STDEV(F2:F35)

	A	B	C	D	E	F	G
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			

工作表1



# 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 $Q_3$ 、下四分位數 $Q_1$ 、  
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

pulipuli.chen@gmail.com

檔案 編輯 檢視 插入 格式 資料 工具 外掛程式 說明 所有變更都已儲存到雲端硬碟 註解 共用

NT\$ % .0 .00 123 連結共用設定已「開啟」。 復原

$\text{fx}$  =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
10		最小值	1	1	1	1	1	1
11		全距	2	3	3	1	2	2
12								
13		上四分位數Q3	2	2	2	2	2	2
14		下四分位數Q1	1	2	2	1	2	1
15		內四分位距	1	0	0	1	0	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								

+ 工作表1



*Thank you for  
your attention*

