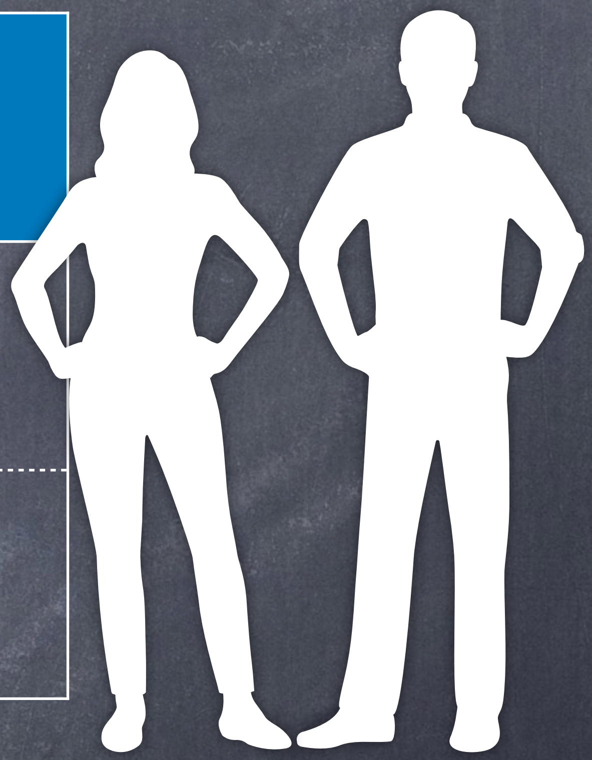
K10 相關係數

數學素養種子與愛思客團隊2021

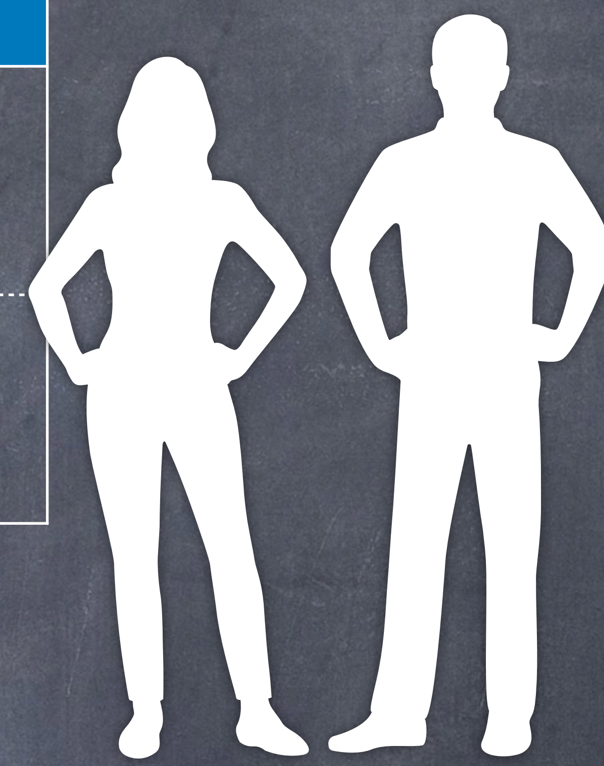
二維數據

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		



二維數據

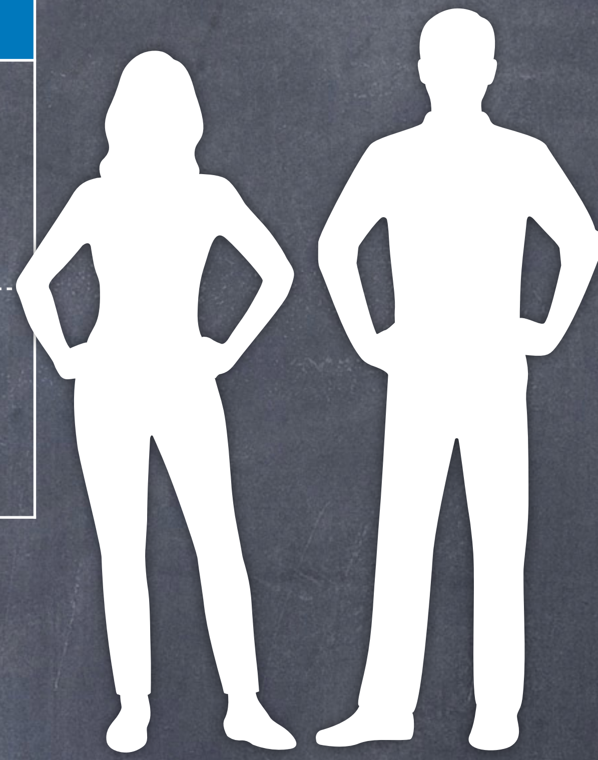
	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		



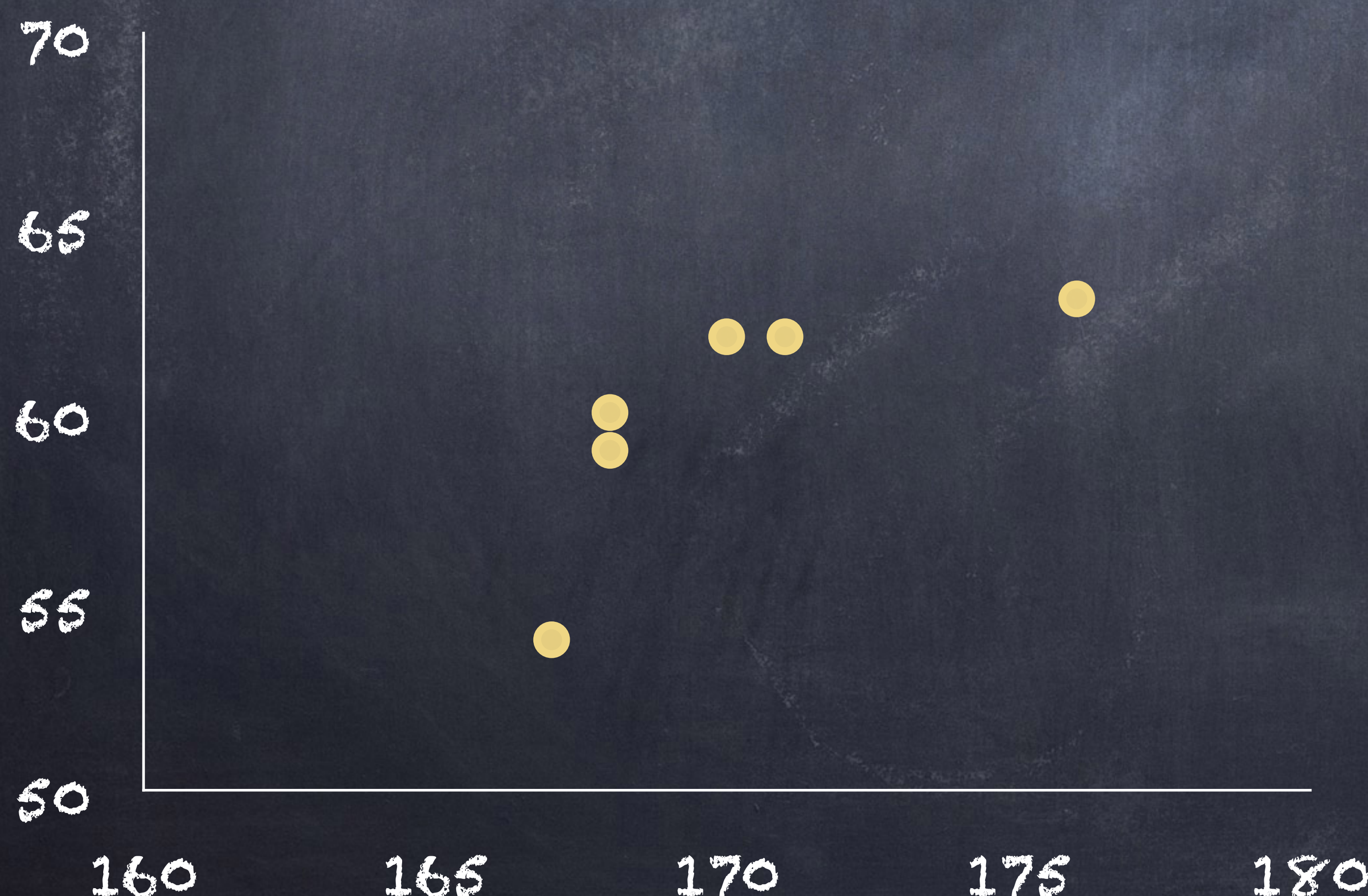
視覺化呈現資料

二維數據

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		



視覺化呈現資料



	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

視覺化呈現資料

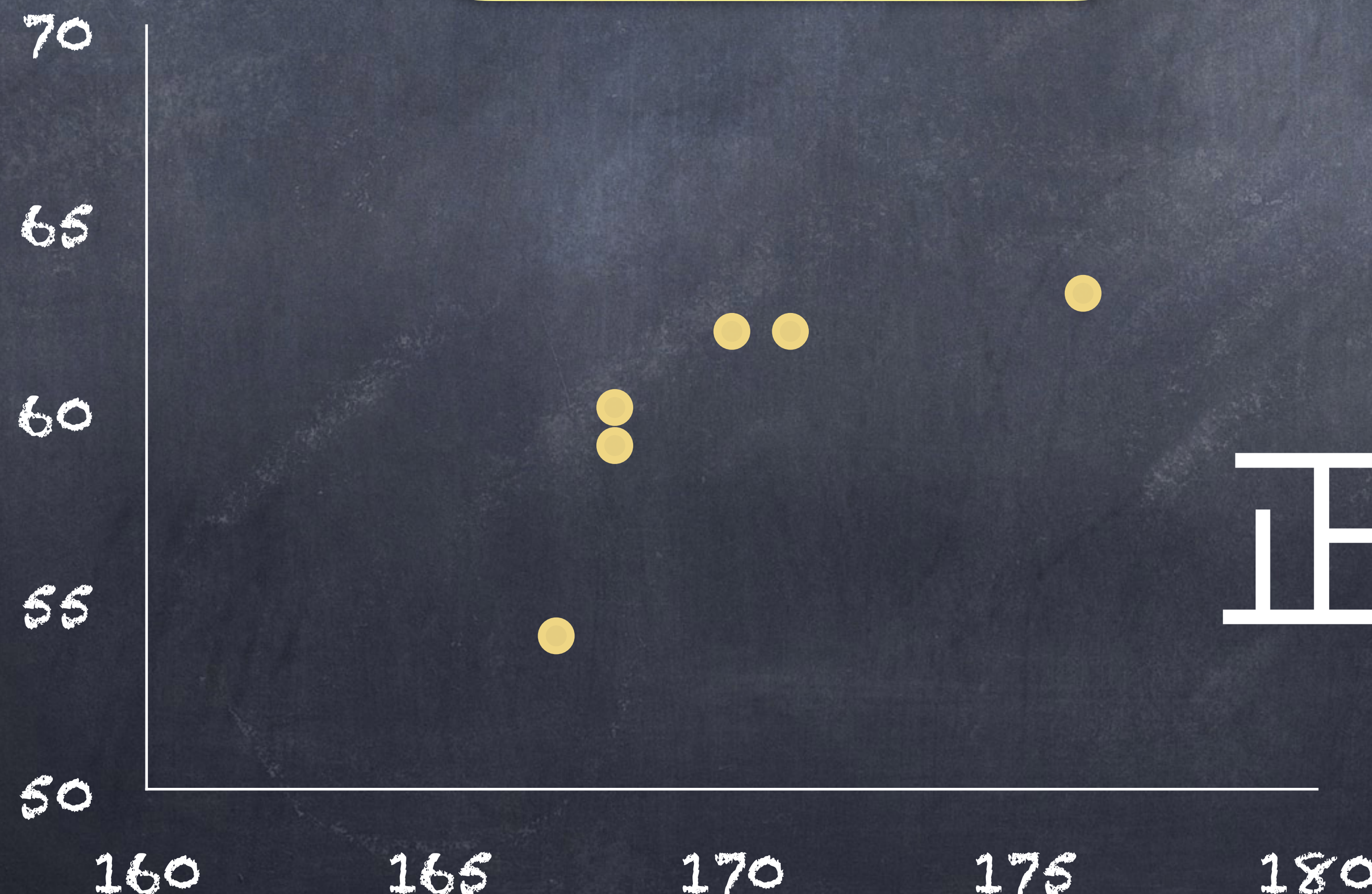
散佈圖

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

散佈圖

視覺化呈現資料



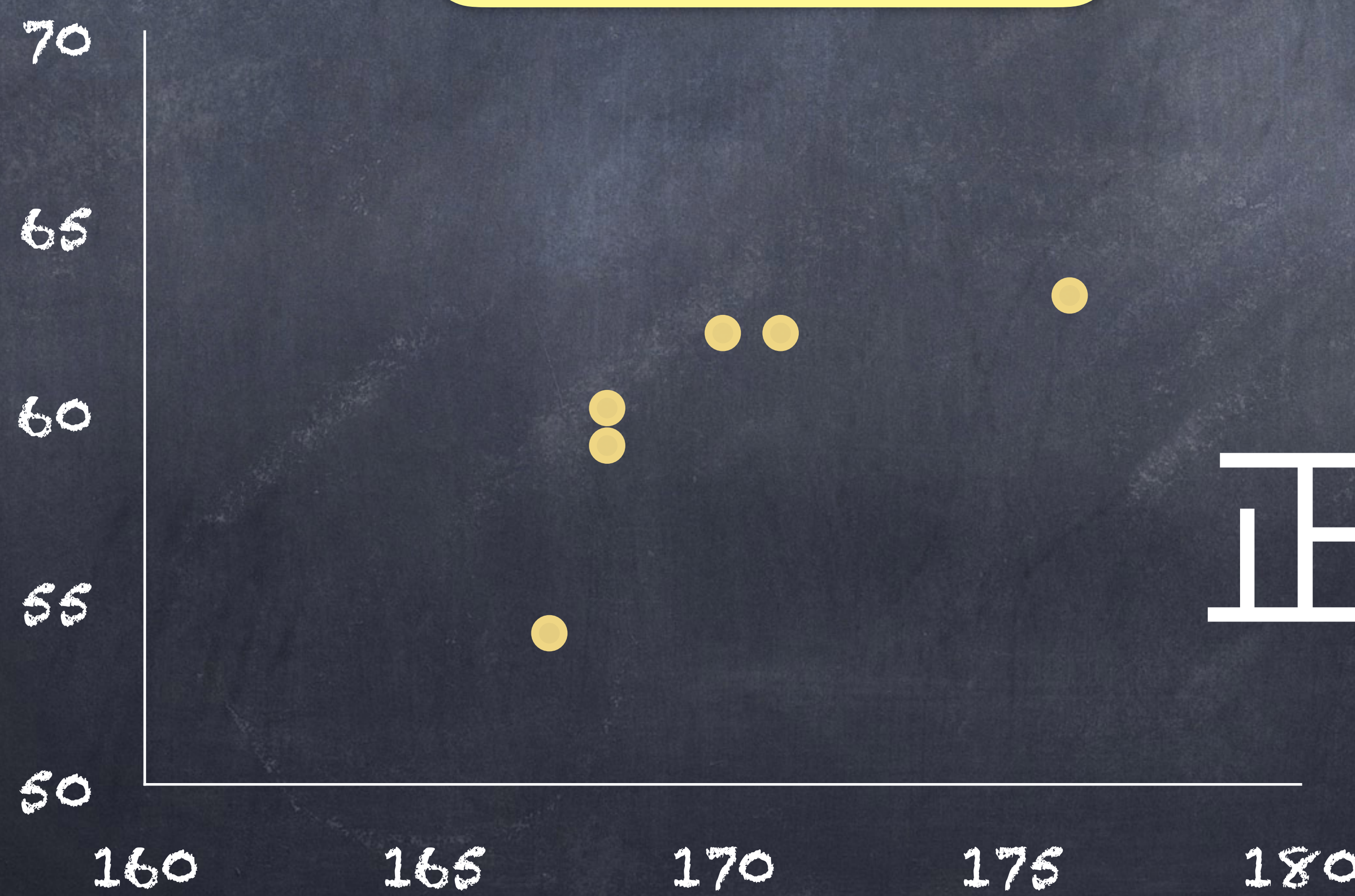
正比？

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

散佈圖

視覺化呈現資料



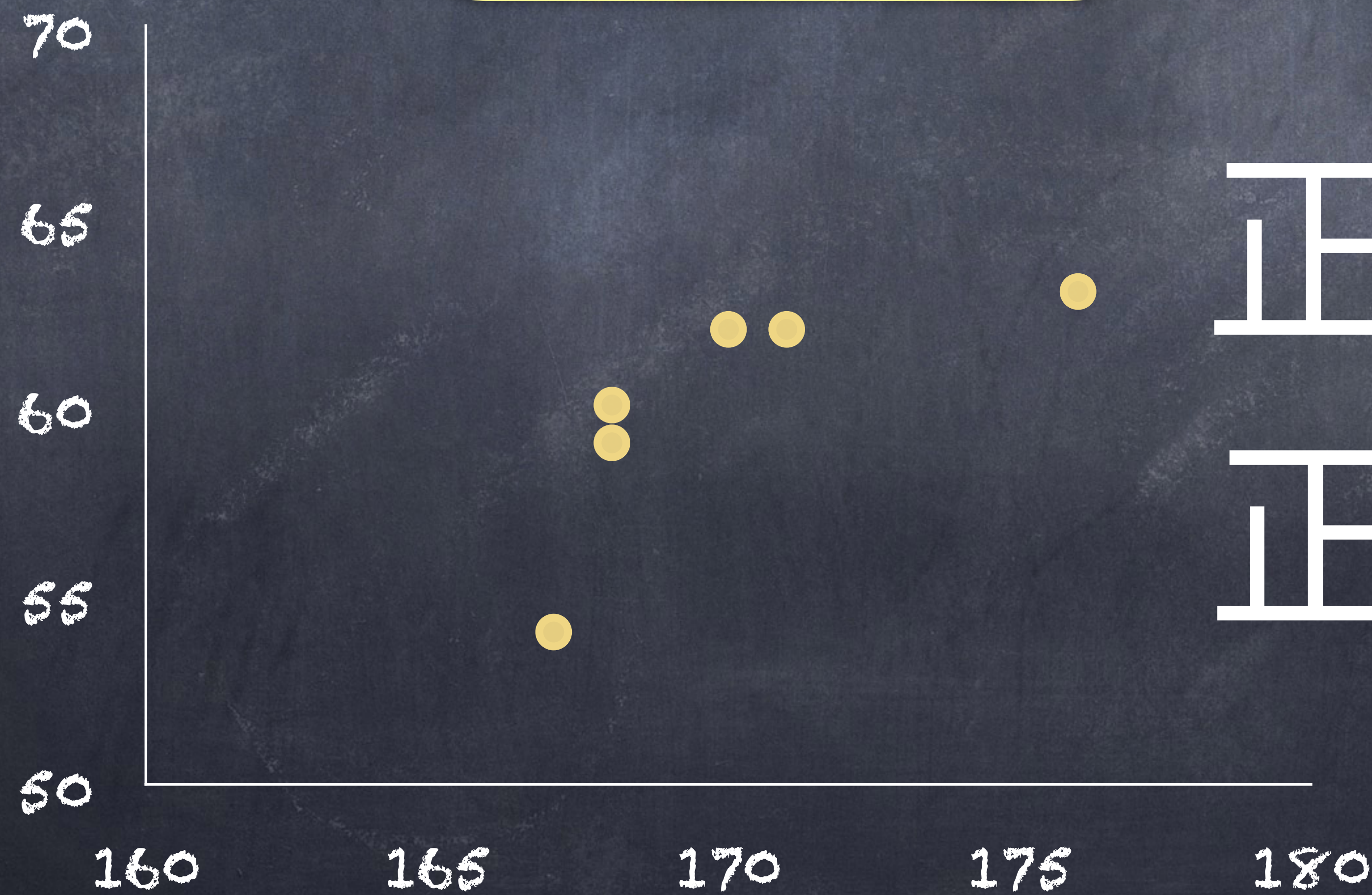
正比

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

散佈圖

視覺化呈現資料



正相關

~~正比~~

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

能不能用一個數值來說明相關程度？

正相關

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

我們需要一個

「新指標」

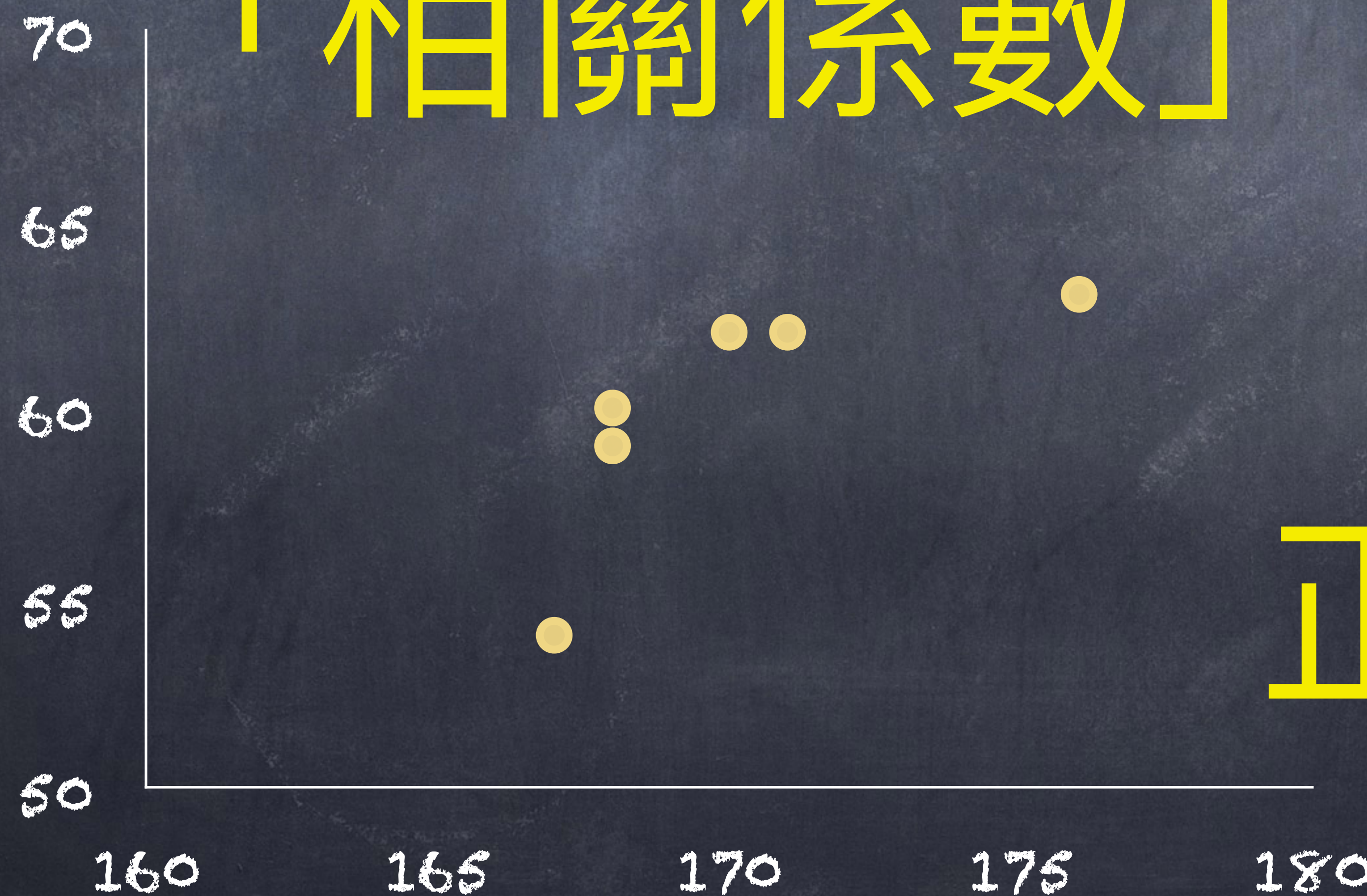
正相關

	A	B	C	D	E	F		
身高	167	168	168	170	171	176		
體重	54	59	60	62	62	63		

二維數據

我們需要

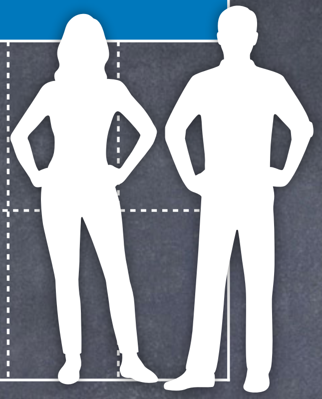
「相關係數」r



正相關

二維數據

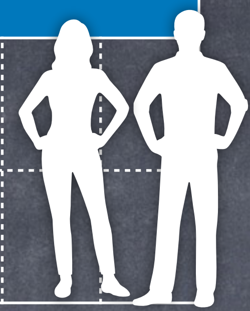
	A	B	C	D	E	F
身高	167	168	168	170	171	176
體重	54	59	60	62	62	63



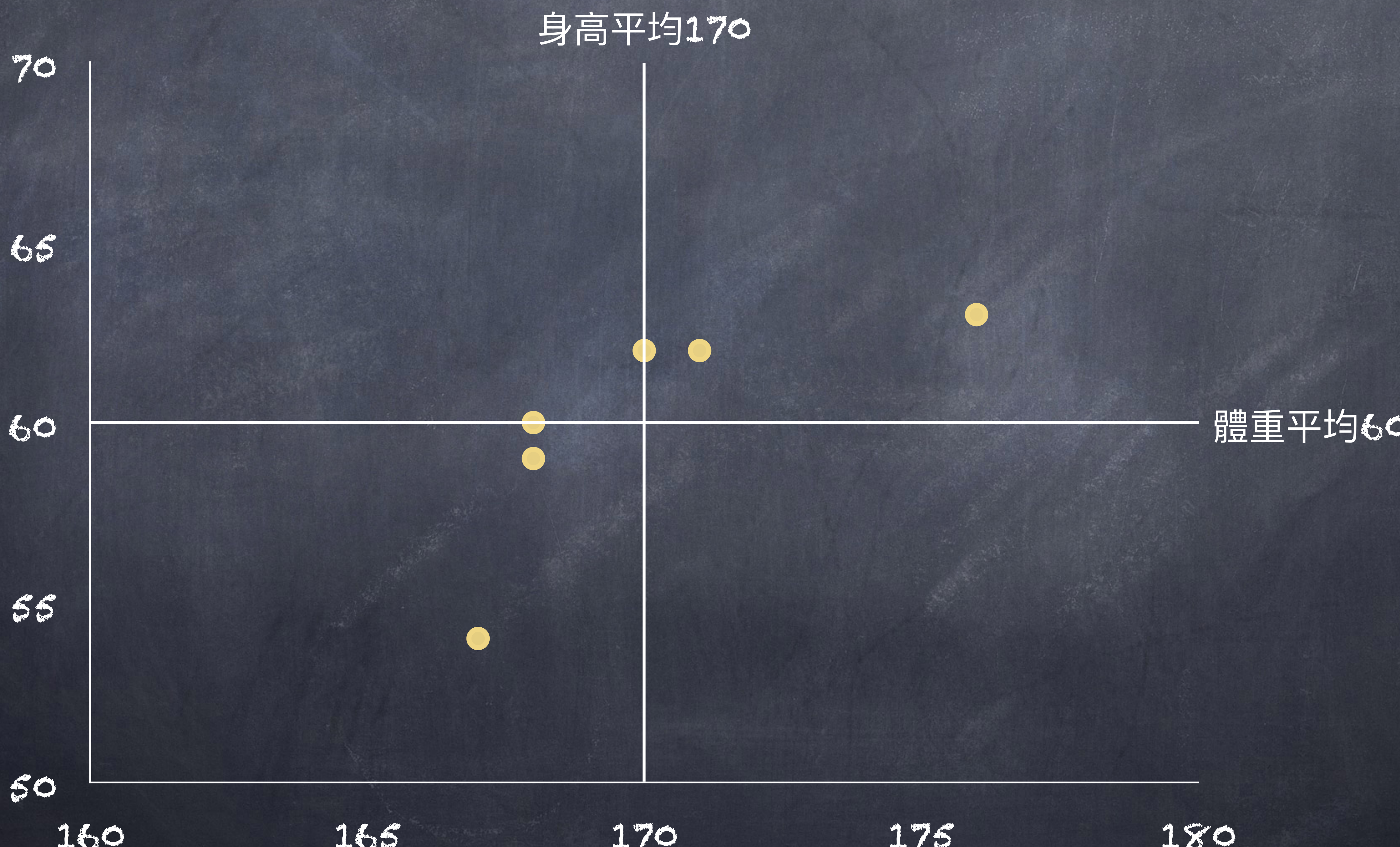
「相關係數」r

二維數據

	A	B	C	D	E	F
身高	167	168	168	170	171	176
體重	54	59	60	62	62	63



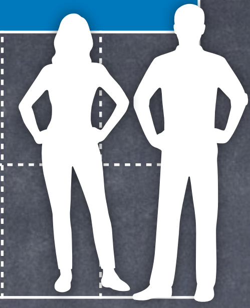
「相關係數」r

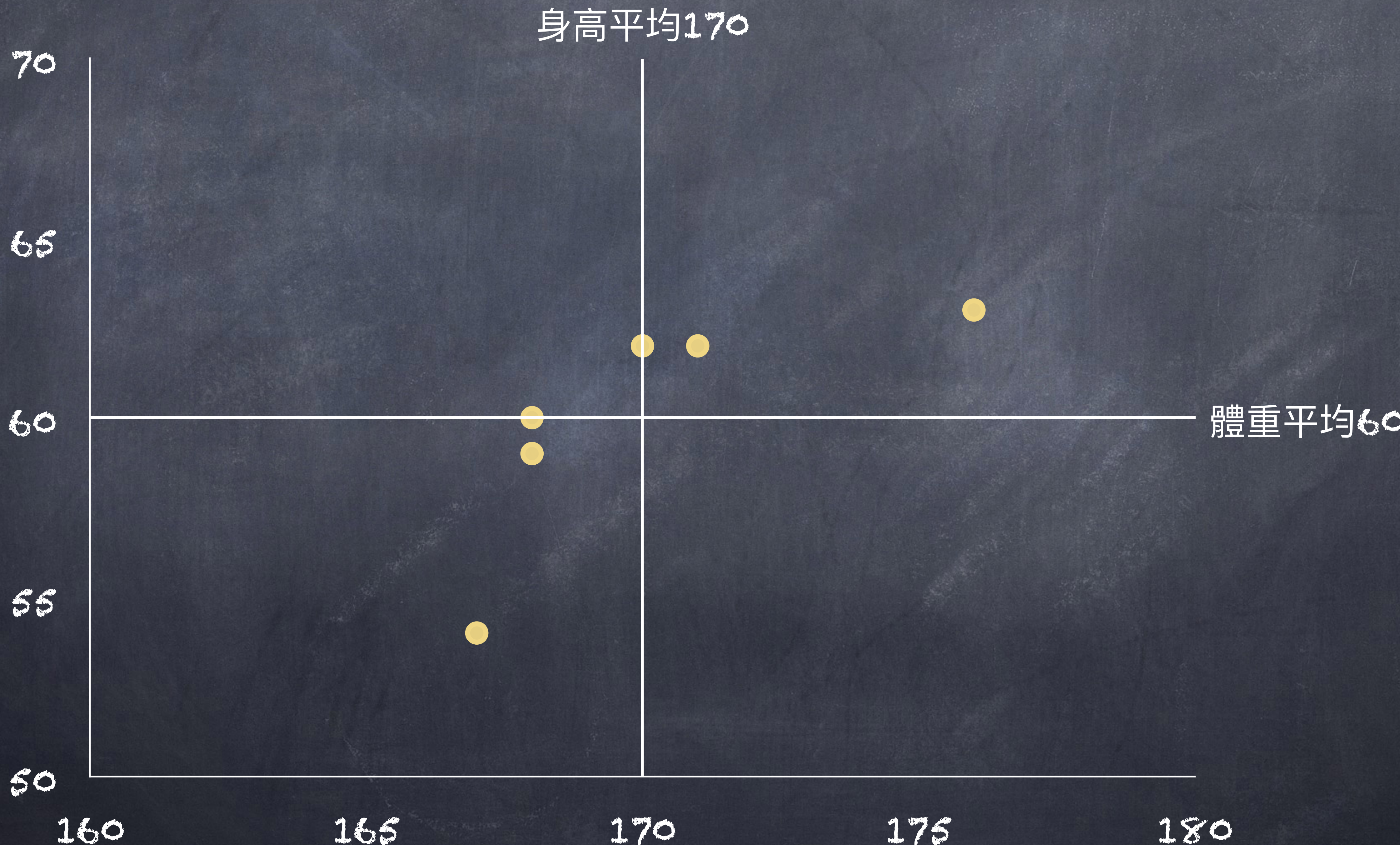


二維數據 「相關係數」r

1.先將數據標準化

	A	B	C	D	E	F
身高	167	168	168	170	171	176
體重	54	59	60	62	62	63





數據標準化

像這種利用伸縮與平移將數據變換，並使得新數據的**平均數為 0**，**標準差為 1** 的方式，稱為**數據標準化**。

數據標準化

設 n 個數據 $x_1, x_2, \cdots, x_n$ 的平均數為 μ，標準差為 σ（$\sigma > 0$）。將每一個數據減去 μ 後再除以 σ，形成一組新數據 $y_1, y_2, \cdots, y_n$，即

$$y_i = \frac{x_i - \mu}{\sigma}, i = 1, 2, \cdots, n。$$

這樣的變換稱為數據標準化。

因為原數據 x_i 的平均數 μ 與標準差 σ 單位相同，所以標準化後的新數據無單位。

標準化數據的運用

	身高	体重	
平均	165	55	
σ	10	5	

μ → 平均

標準差 → σ

170. 60

$$\frac{x-\mu}{\sigma}$$

標準化

身 $\frac{170-165}{10}=\frac{1}{2}$

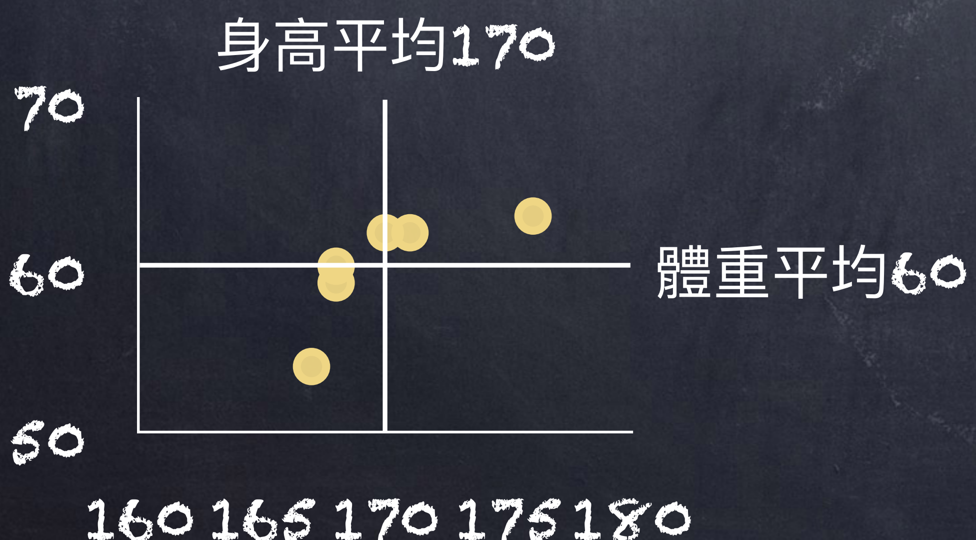
体 $\frac{60-55}{5}=1$

体重發育好

二維數據 「相關係數」r

1.先將數據標準化

	A	B	C	D	E	F	平均	標準差
身高	167	168	168	170	171	176	170	3
體重	54	59	60	62	62	63	60	3



二維數據

「相關係數」r

1.先將數據標準化

	A	B	C	D	E	F	平均	標準差
身高	167	168	168	170	171	176	170	3
體重	54	59	60	62	62	63	60	3

x

$\Downarrow$

平均

$$\frac{x-\mu}{\sigma}$$

標準差

	A	B	C	D	E	F	平均	標準差
身高	−1	−2/3	−2/3	0	1/3	2	0	1
體重	−2	−1/3	0	2/3	2/3	1	0	1

二維數據　「相關係數」r

1.先將數據標準化

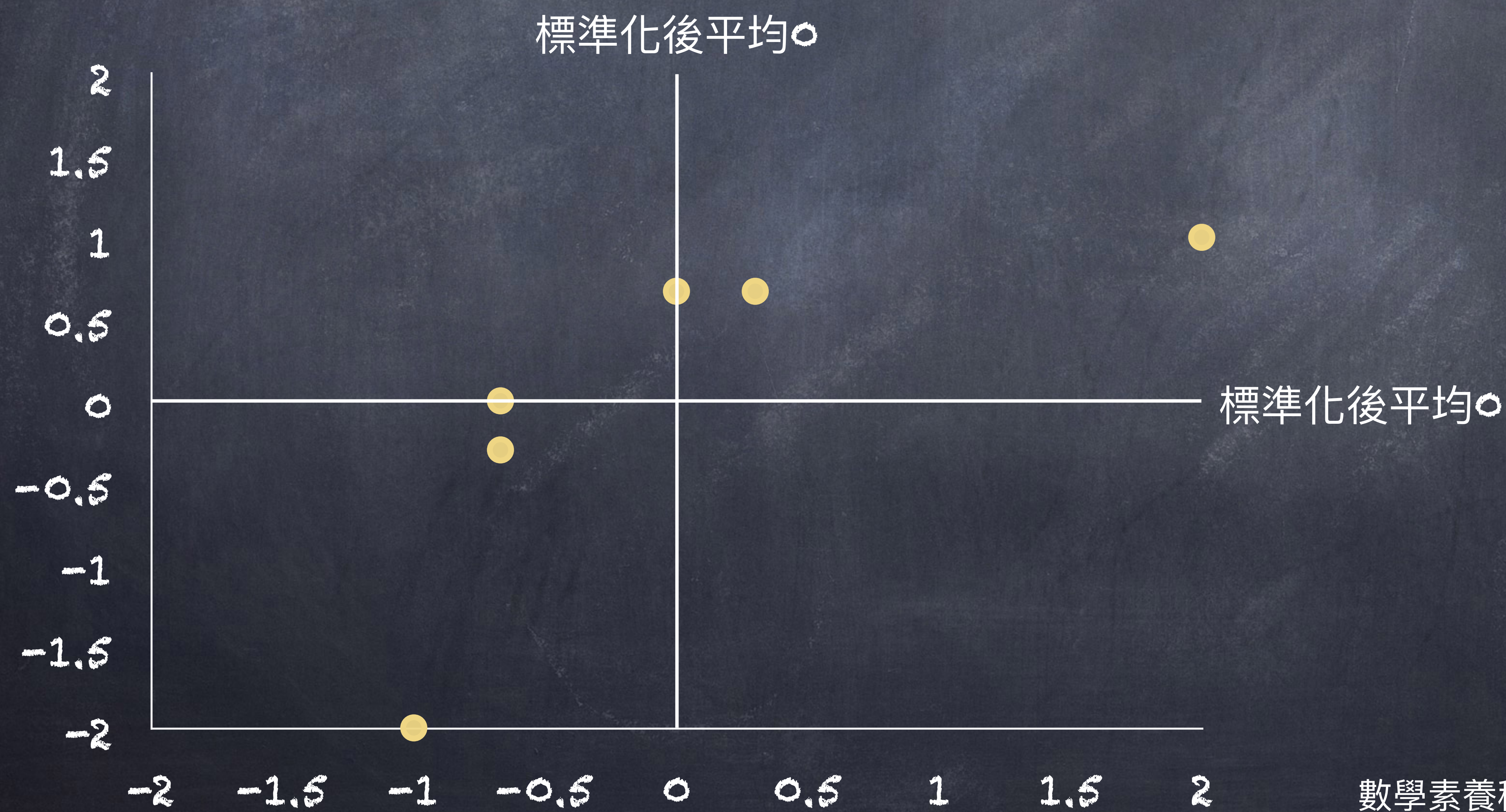
	A	B	C	D	E	F	平均	標準差
身高	−1	−2/3	−2/3	0	1/3	2	0	1
體重	−2	−1/3	0	2/3	2/3	1	0	1

二維數據

「相關係數」r

	A	B	C	D	E	F	平均	標準差
身高	−1	−2/3	−2/3	0	1/3	2	0	1
體重	−2	−1/3	0	2/3	2/3	1	0	1

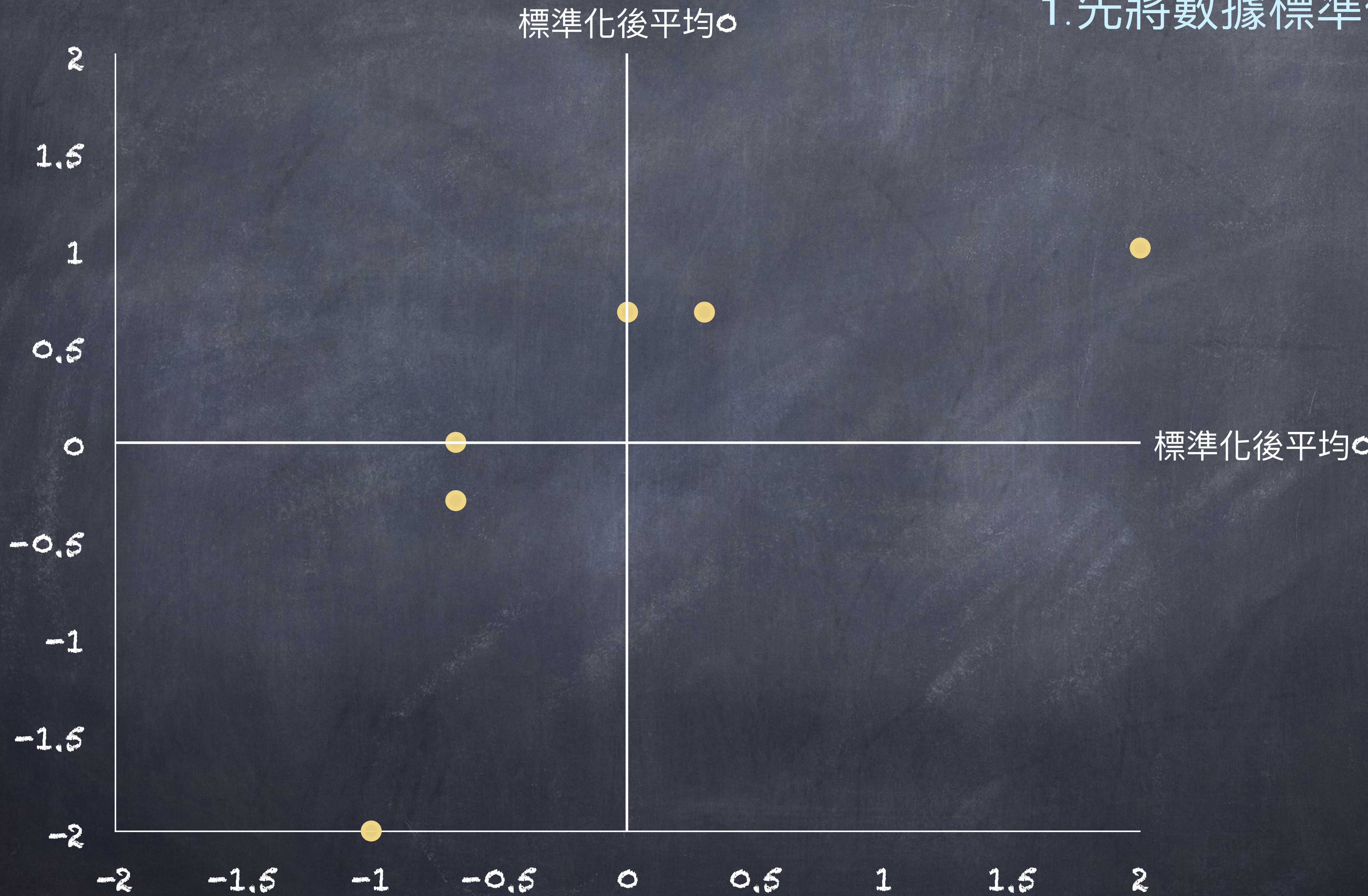
1.先將數據標準化



二維數據

「相關係數」r

1.先將數據標準化

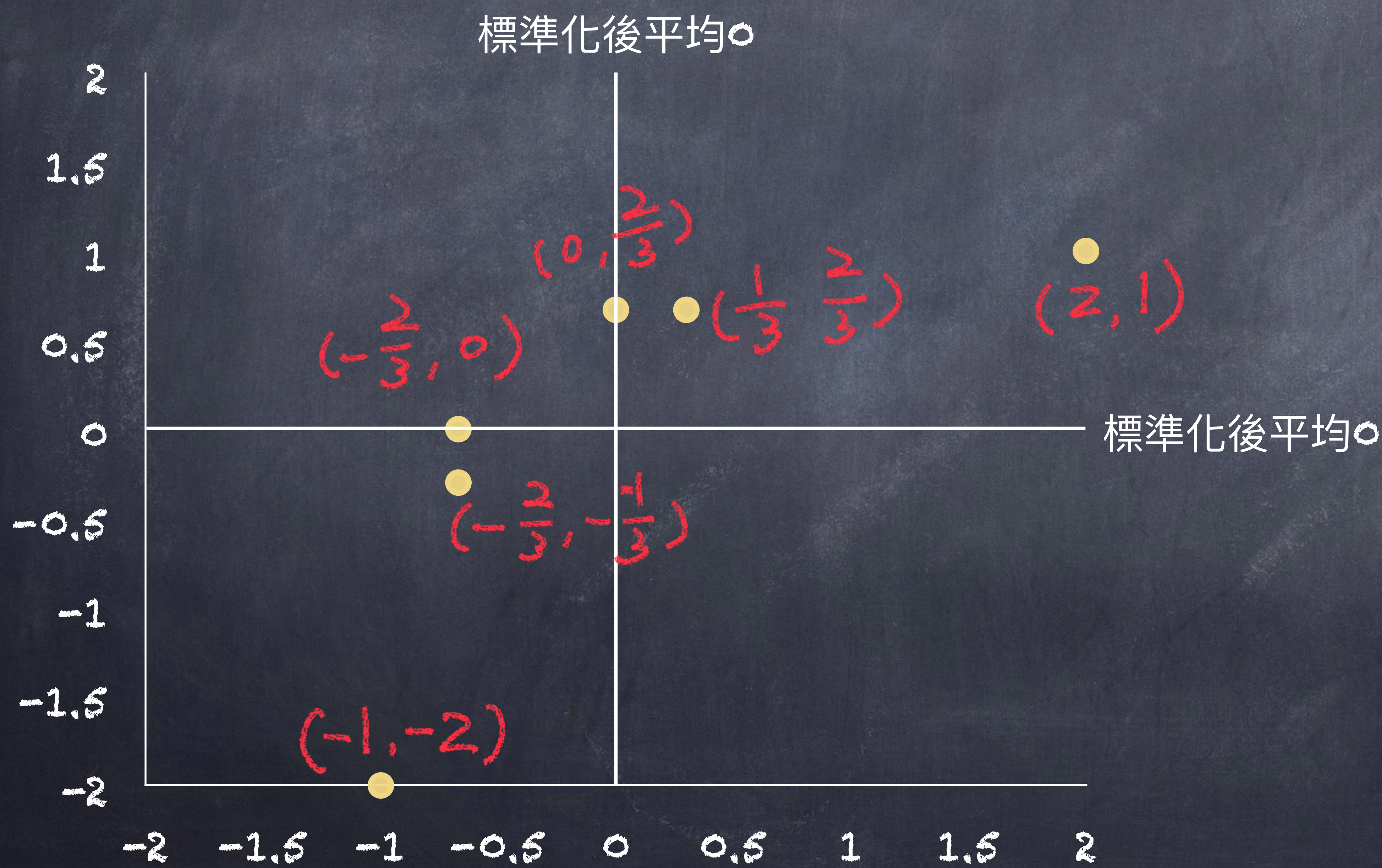


「相關係數」r

二維數據

1. 先將數據標準化

	A	B	C	D	E	F	平均	標準差
身高	−1	−2/3	−2/3	0	1/3	2	0	1
體重	−2	−1/3	0	2/3	2/3	1	0	1



「相關係數」r

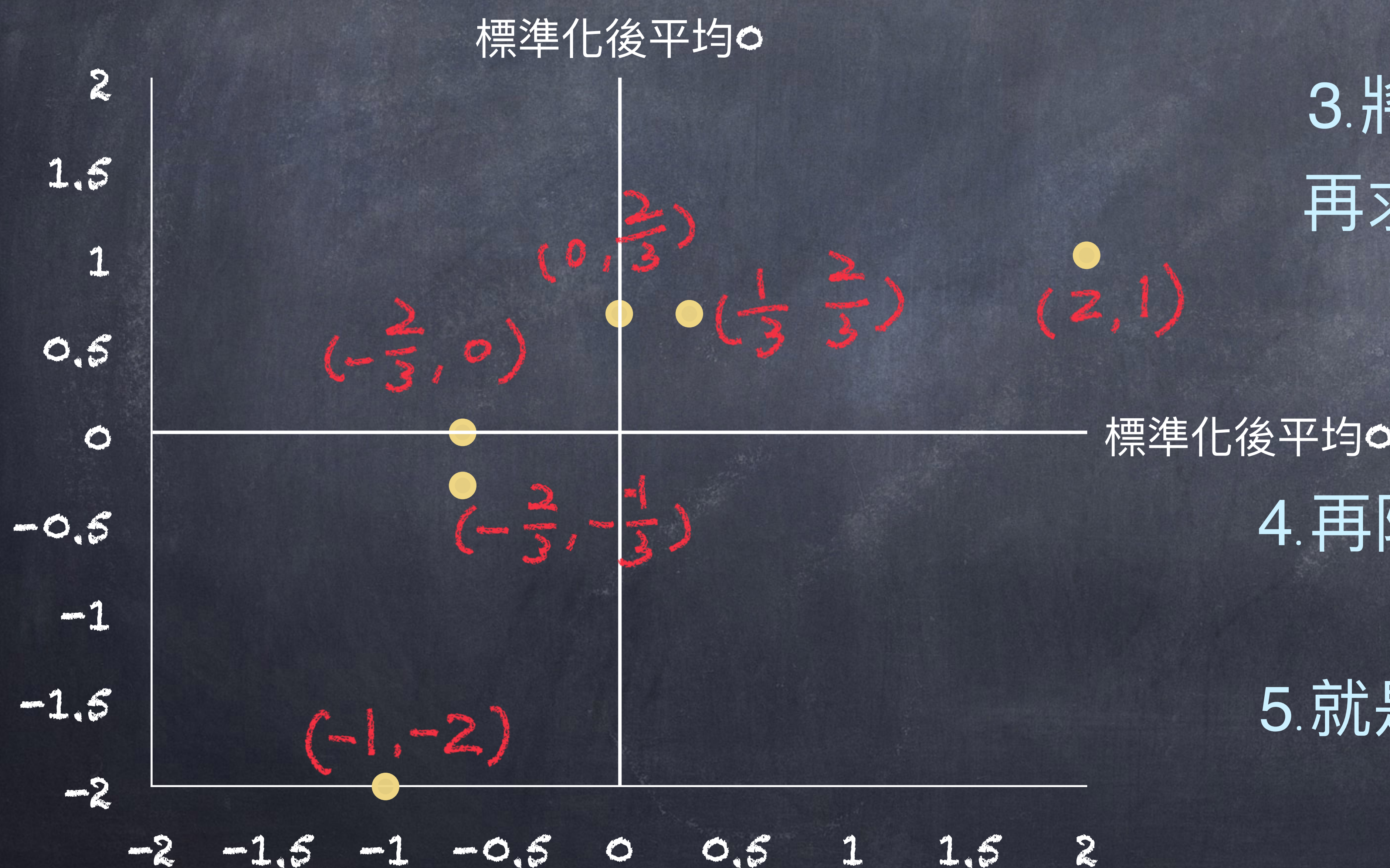
二維數據

	A	B	C	D	E	F	平均	標準差
身高	—1	—2/3	—2/3	0	1/3	2	0	1
體重	—2	—1/3	0	2/3	2/3	1	0	1

1. 先將數據標準化
2. 因++得+

   － －得+
2. 因+ －得－

   － +得－
3. 將數據標準化相乘再求總和，可知正、負、零相關性
4. 再除以n，抵銷數量造成的影響
5. 就是相關係數 r

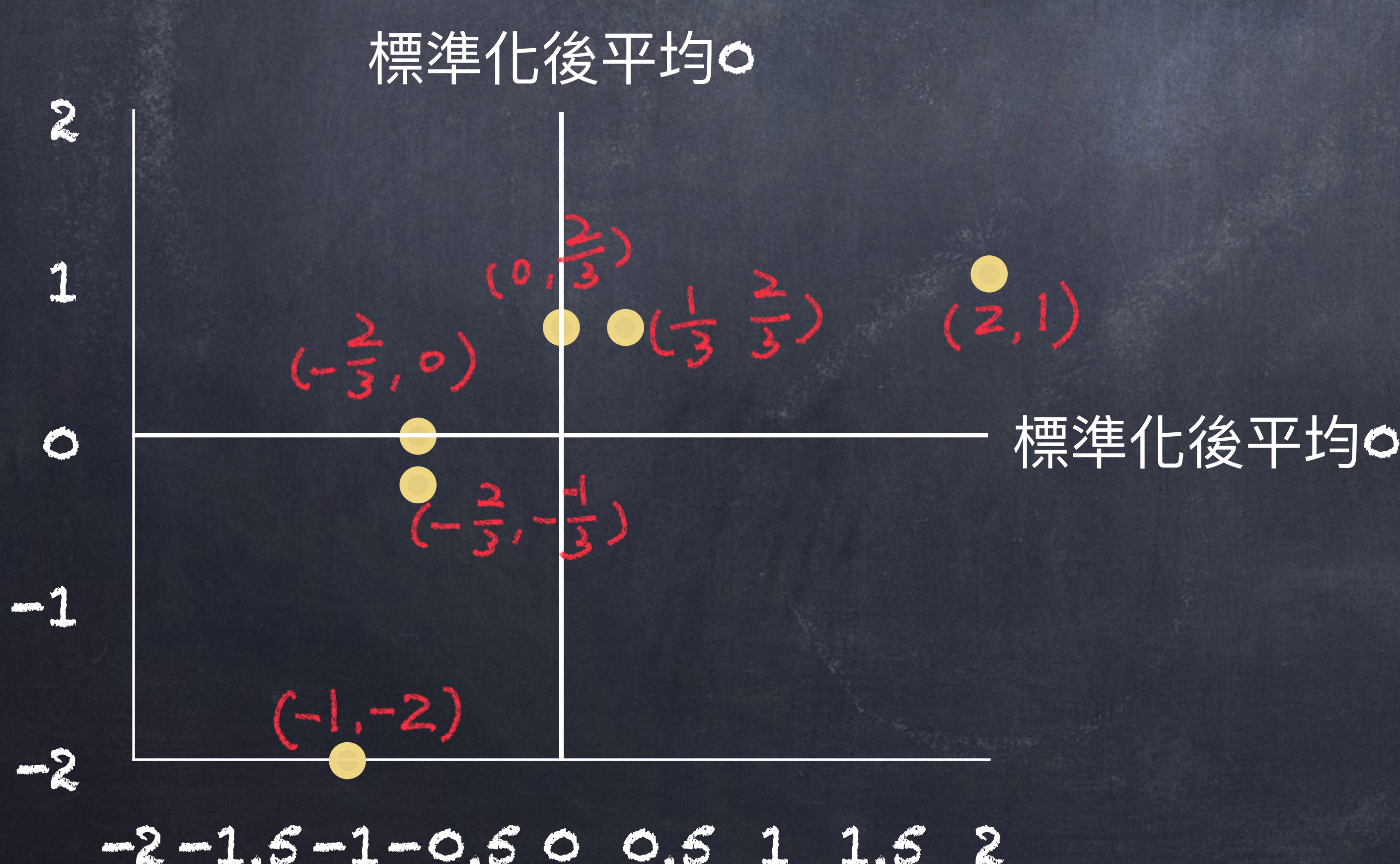


二維數據

「相關係數」r

	A	B	C	D	E	F	平均	標準差
身高	−1	−2/3	−2/3	0	1/3	2	0	1
體重	−2	−1/3	0	2/3	2/3	1	0	1
相乘	2	4/9	0	0	2/9	2	總和	24/9
							除以n	

1. 先將數據標準化
2. 因++得+
   − −得+
2. 因+ −得−
   − +得−
3. 將數據標準化相乘再求總和，可知正、負、零相關性
4. 再除以n，抵銷數量造成的影響
5. 就是相關係數 r



二維數據

	A	B	C	D	E	F	平均	標準差
身高	−1	−2/3	−2/3	0	1/3	2	0	1
體重	−2	−1/3	0	2/3	2/3	1	0	1
相乘	2	2/9	0	0	2/9	2	總和	40/9
							除以n	40/54

「相關係數」r

1. 先將數據標準化
2. 因++得+

   −−得+

2. 因+−得−

   −+得−

3. 將數據標準化相乘再求總和，可知正、負、零相關性
4. 再除以n，抵銷數量造成的影響
5. 就是相關係數 r

相關係數

$r = 40/54$

例題 2

$(1+2+3+4+5)\div5=3$

$\sigma=\sqrt{\frac{1+4+9+16+25}{5}-9}=\sqrt{2}$ 答對

兩變量 x 與 y 的數據如下表。

x	1	2	3	4	5	平均 3	標準差 $\sqrt{2}$
y	4	5	3	1	2	3	$\sqrt{2}$

(1) 繪出 x 與 y 的散布圖。

(2) 求 x 與 y 的相關係數。

$\frac{x-\mu}{\sigma} \Leftrightarrow \frac{資料-平均}{標準差}$

解 ①數據標準化 ②相乘再全部相加

x	1	2	3	4	5	
y	4	5	3	1	2	
x'	$\frac{-2}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	0	$\frac{1}{\sqrt{2}}$	$\frac{2}{\sqrt{2}}$	相加
y'	$\frac{1}{\sqrt{2}}$	$\frac{2}{\sqrt{2}}$	0	$\frac{-2}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	
$x'y'$	$\frac{-2}{2}$	$\frac{-2}{2}$	0	$\frac{-2}{2}$	$\frac{-2}{2}$	-4

相乘

$n=5$

③除以 n

$$\boxed{\frac{-4}{5}}$$

$$r=\left[\frac{(x_1-\mu_x)}{\sigma_x}\times\frac{(y_1-\mu_y)}{\sigma_y}+\frac{(x_2-\mu_x)}{\sigma_x}\times\frac{(y_2-\mu_y)}{\sigma_y}+\cdots\cdots\right]/n$$

$$=\frac{1}{n}\left[\frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots\cdots}{\sigma_x\times\sigma_y}\right]$$

故 $r=\frac{-4}{5}$

$r=-0.8$ #

單元09-二維數據分析-教師用書 p...

例題 2

$(1+2+3+4+5)\div5=3$

$\sigma=\sqrt{\frac{1+4+9+16+25}{5}-9}=\sqrt{2}$ 答對

兩變量 x 與 y 的數據如下表。

x	1	2	3	4	5	平均 3	標準差 $\sqrt{2}$
y	4	5	3	1	2	3	$\sqrt{2}$

(1) 繪出 x 與 y 的散布圖。

(2) 求 x 與 y 的相關係數。

$\frac{x-\mu}{\sigma} \Leftrightarrow \frac{資料-平均}{標準差}$

解 ①數據標準化 ②相乘再全部相加

x	1	2	3	4	5	
y	4	5	3	1	2	
x'	$\frac{-2}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	0	$\frac{1}{\sqrt{2}}$	$\frac{2}{\sqrt{2}}$	相加
y'	$\frac{1}{\sqrt{2}}$	$\frac{2}{\sqrt{2}}$	0	$\frac{-2}{\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	
$x'y'$	$\frac{-2}{2}$	$\frac{-2}{2}$	0	$\frac{-2}{2}$	$\frac{-2}{2}$	-4

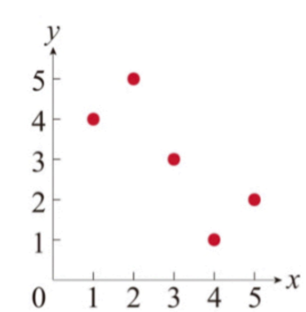
相乘

$n=5$

③除以 n

$$\boxed{\frac{-4}{5}}$$

x 與 y 的散布圖如下圖。



故 $r=-0.8$ #

(相關係數 $=-0.8$)

a. ⊖ ⇒ 負相關

b. $|-0.8|=0.8$ ⇒

表(?)值量大小

web

$$r=\left[\frac{(x_1-\mu_x)}{\sigma_x}\times\frac{(y_1-\mu_y)}{\sigma_y}+\frac{(x_2-\mu_x)}{\sigma_x}\times\frac{(y_2-\mu_y)}{\sigma_y}+\cdots\cdots\right]/n$$

$$=\frac{1}{n}\left[\frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots\cdots}{\sigma_x\times\sigma_y}\right]$$

$$=\frac{1}{n}\left[\frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots\cdots}{\sqrt{\frac{1}{n}\left[(x_1-\mu_x)^2+(x_2-\mu_x)^2+\cdots\right]}\sqrt{\frac{1}{n}\left[(y_1-\mu_y)^2+(y_2-\mu_y)^2+\cdots\right]}}\right]$$

① 數據標準化, $\frac{x-\mu}{\sigma}$

② 相乘再全相加

③ 除以 n

$\Rightarrow$ r（相關係數）

標準差公式

設 n 個數據 $x_1,x_2,\cdots,x_n$ 的平均數為 μ，則標準差為

$$\sigma\overset{①}{=}\sqrt{\frac{1}{n}\left((x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2\right)}$$

$$\overset{②}{=}\sqrt{\frac{1}{n}\left(x_1^2+x_2^2+\cdots+x_n^2\right)-\mu^2}\ 。$$

$$r=\frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots\cdots}{\sqrt{(x_1-\mu_x)^2+(x_2-\mu_x)^2+\cdots\cdots}\sqrt{(y_1-\mu_y)^2+(y_2-\mu_y)^2+\cdots\cdots}}$$

相關係數的定義

設兩變量 x 與 y 的 n 筆數據為 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，且 x 與 y 的平均數分別為 μ_x,μ_y，標準差分別為 σ_x,σ_y。定義兩變量 x 與 y 的相關係數為

① $$r=\frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots+(x_n-\mu_x)(y_n-\mu_y)}{n\sigma_x\sigma_y}\ 。$$

$$=\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

以下我們導出相關係數的另一公式，令

$$S_{xy}=(x_1-\mu_x)(y_1-\mu_y)+\cdots+(x_n-\mu_x)(y_n-\mu_y),$$

$$S_{xx}=(x_1-\mu_x)^2+\cdots+(x_n-\mu_x)^2,$$

$$S_{yy}=(y_1-\mu_y)^2+\cdots+(y_n-\mu_y)^2,$$

則標準差 σ_x 與 σ_y 可改寫為

$$\sigma_x=\sqrt{\frac{1}{n}\left((x_1-\mu_x)^2+\cdots+(x_n-\mu_x)^2\right)}=\sqrt{\frac{S_{xx}}{n}},$$

$$\sigma_y=\sqrt{\frac{1}{n}\left((y_1-\mu_y)^2+\cdots+(y_n-\mu_y)^2\right)}=\sqrt{\frac{S_{yy}}{n}}\ 。$$

相關係數的計算公式

設兩變量 x 與 y 的 n 筆數據為 $(x_1,y_1),(x_2,y_2),\ \cdots,\ (x_n,y_n)$，其相關係數

② $$r=\frac{S_{xy}}{\sqrt{S_{xx}}\cdot\sqrt{S_{yy}}}\ 。$$

a. $(x_i-\mu_x)$

b. $(y_i-\mu_y)$

c. $(x_i-\mu_x)^2$

d. $(y_i-\mu_y)^2$

e. $(x_i-\mu_x)(y_i-\mu_y)$ 的和

首頁
分類索引
特色內容
新聞動態
近期變更
隨機條目
資助維基百科

說明
說明
維基社群
方針與指引
互助客棧
知識問答
字詞轉換
IRC即時聊天
聯絡我們
關於維基百科

工具
連結至此的頁面
相關變更
上傳檔案
特殊頁面
固定連結
頁面資訊
引用此頁面
維基數據項目
短網址

列印/匯出
下載為 PDF
可列印版

皮爾森積動差相關係數 [編輯]

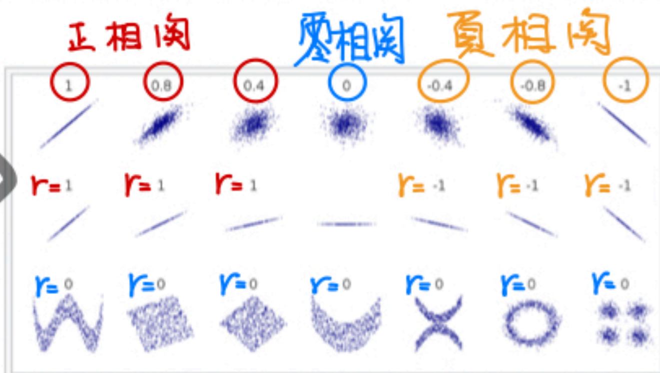
$-1 \leq r \leq 1$

維基百科，自由的百科全書

> 此條目翻譯品質不佳。（2017年4月4日）
> 翻譯者可能不熟悉中文或原文語言，也可能使用了機器翻譯，請協助翻譯本條目或重新編寫，並注意避免翻譯腔的問題。明顯拙劣的機器翻譯請改掛 {{d|G13}} 提交刪除。

在統計學中，**皮爾森積動差相關係數**（英語：Pearson product-moment correlation coefficient，又稱作 PPMCC或PCCs[1], 文章中常用r或Pearson's r表示）用於度量兩個變數X和Y之間的相關程度（線性相依），其值介於–1與1之間。在自然科學領域中，該係數廣泛用於度量兩個變數之間的線性相依程度。它是由卡爾·皮爾森從弗朗西斯·高爾頓在19世紀80年代提出的一個相似卻又稍有不同的想法演變而來。[2][3]這個相關係數也稱作「皮爾森相關係數r」。

目錄 [隱藏]



幾組(x, y)的點集，以及各個點集中x和y之間的相關係數。我們可以發現相關係數反映的是變數之間的線性關係和相關性的方向（第一排），而不是相關性的斜率（中間），也不是各種非線性關係（第三排）。請注意：中間的圖中斜率為0，但相關係數是沒有意義的，因為此時變數Y是0

$$r = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}}$$

$$r = \frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots\cdots}{\sqrt{(x_1-\mu_x)^2+(x_2-\mu_x)^2+\cdots\cdots}\sqrt{(y_1-\mu_y)^2+(y_2-\mu_y)^2+\cdots\cdots}}$$

$$-1 \leq r \leq 1$$

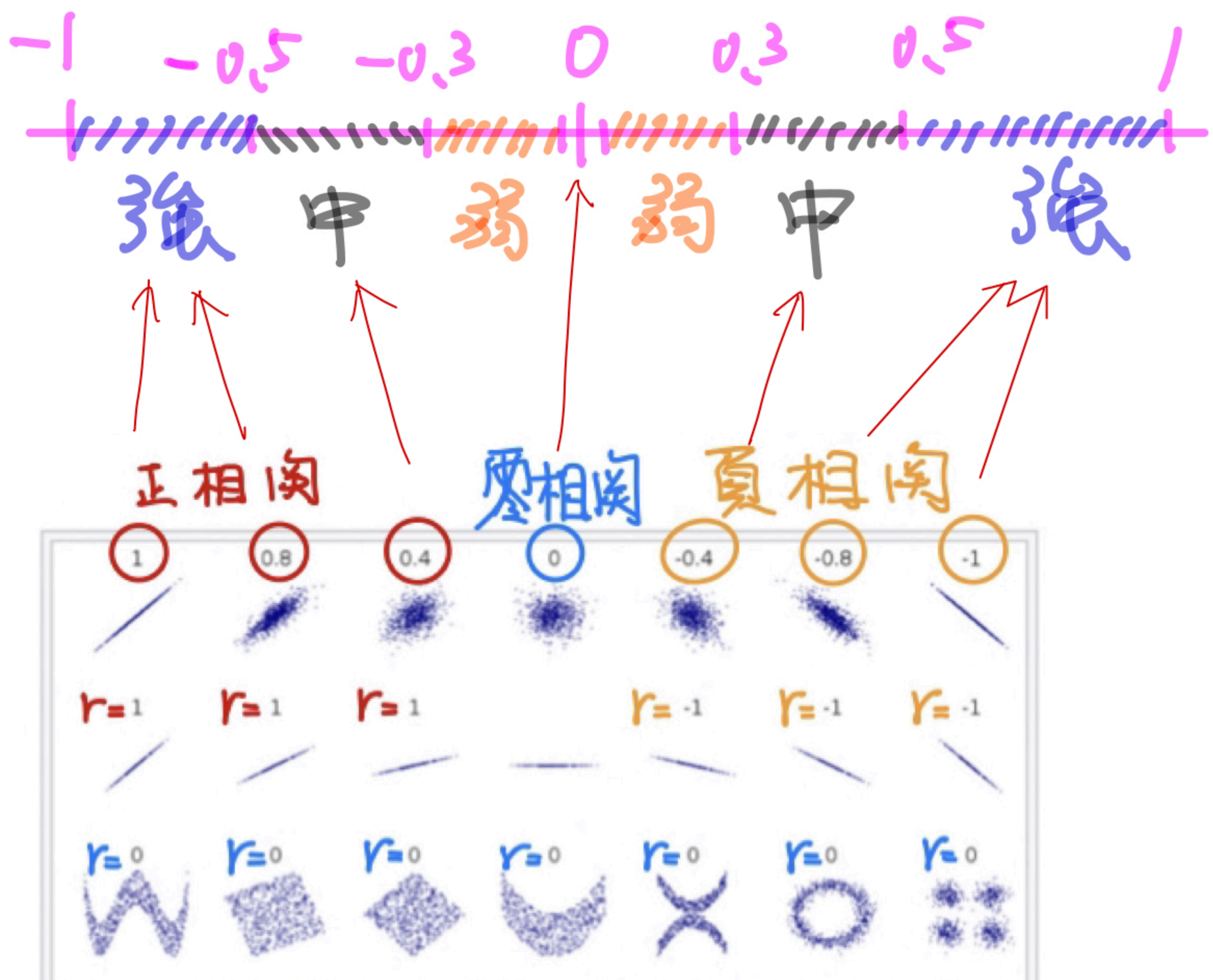
相關性	負	正
無	−0.09 to 0.0	0.0 to 0.09
弱	−0.3 to −0.1	0.1 to 0.3
中	−0.5 to −0.3	0.3 to 0.5
強	−1.0 to −0.5	0.5 to 1.0

$0 \leq |r| \leq 0.09$ 幾乎無相關

$0.1 \leq |r| \leq 0.3$ 弱相關

$0.3 \leq |r| \leq 0.5$ 中相關

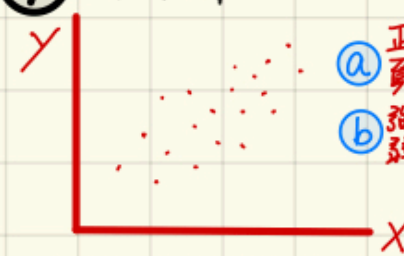
$0.5 \leq |r| \leq 1$ 強相關



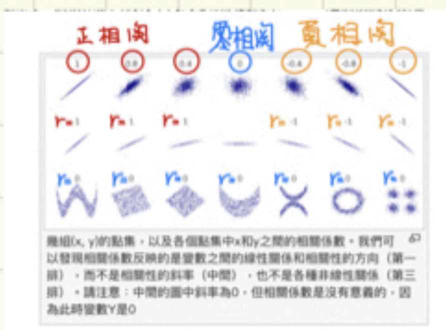
幾組(x, y)的點集，以及各個點集中x和y之間的相關係數。我們可以發現相關係數反映的是變數之間的線性關係和相關性的方向（第一排），而不是相關性的斜率（中間），也不是各種非線性關係（第三排）。請注意：中間的圖中斜率為0，但相關係數是沒有意義的，因為此時變數Y是0

單元09-二維數據分析-教師用書 pdf 檔

① 散佈圖



ⓐ 正/負 趨勢性
ⓑ 強/弱 關連性

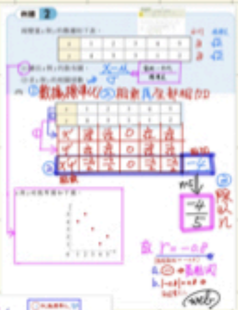


② 相關係數（用一個數，未說明 ⓐ 與 ⓑ

r = ① 數據標準化 $\frac{x-\mu}{\sigma}$
② $x'y'$ 相乘再全部相加
③ 除以 n，即得到 r

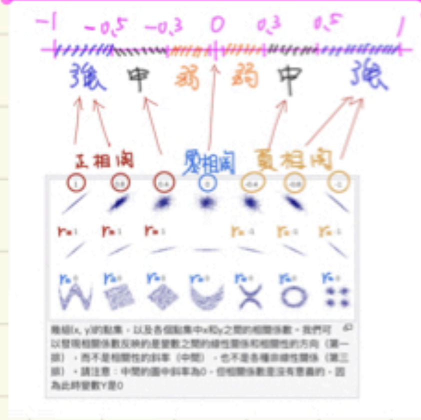
$$r=\frac{(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots\cdots}{\sqrt{(x_1-\mu_x)^2+(x_2-\mu_x)^2+\cdots\cdots}\sqrt{(y_1-\mu_y)^2+(y_2-\mu_y)^2+\cdots\cdots}}$$

$$r=\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$



相關性	負	正
無	−0.09 to 0.0	0.0 to 0.09
弱	−0.3 to −0.1	0.1 to 0.3
中	−0.5 to −0.3	0.3 to 0.5
強	−1.0 to −0.5	0.5 to 1.0

$0\le|r|\le0.09$ 幾乎無相關
$0.1\le|r|\le0.3$ 弱相關
$0.3\le|r|\le0.5$ 中相關
$0.5\le|r|\le1$ 強相關



r 說明了 ⓐ ⓑ

$-1\le r\le1$

r ⊕/⊖ ⇒ ⓐ

$|r|$ 大/小 ⇒ ⓑ

r 沒有單位

05/27作業

相關係數是 ____________________

相關係數不是 ____________________