二維數據分析【基礎篇】

【主題 1】散布圖與相關係數

焦點 1. 散布圖

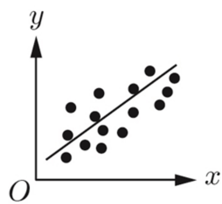
(1) 如果要考量兩個變數之間的關聯，那麼就需要搜集成對的變數資料，也就是二維數據。而想要呈現二維數據所含的訊息，最常用的圖形就是散布圖。

(2) 假設我們對 n 個個體蒐集了兩個變數 X, Y 的資料，第一個變數當作 x 坐標，第二個變數當作 y 坐標，用 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 表示。將這 n 個數據視為 xy 坐標平面上的點，描繪在 xy 坐標平面上，就構成了 X, Y 兩個變數的散布圖。

焦點 2. 相關的分類

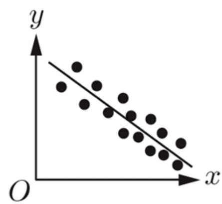
(1) 正相關

當散布圖中的各點，大致上看來在一條直線的附近時，我們就稱變數 X 和 Y 具有直線關聯。當 X, Y 散布圖中的點大致呈現從左下方到右上方的趨勢時，這是變數 X 和 Y 為正相關的概念。



(2) 負相關

當 X, Y 散布圖中的點大致呈現從左上方到右下方的趨勢時，這是變數 X 和 Y 為負相關的概念。



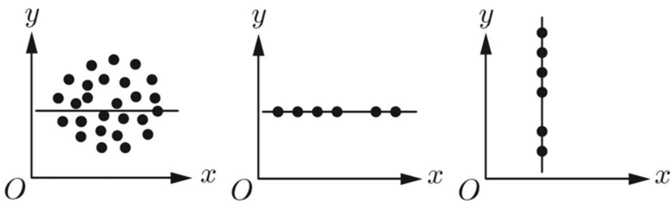
(3) 零相關

如果點的散布情形，沒有辦法觀察出任何帶狀的形式，所呈現的狀況，代表 X 和 Y 幾乎沒有直線關聯。

散布圖的點呈現水平線的分布，即無論 x 的值為何，y 的值均為定值。

散布圖的點呈現鉛直線的分布，即無論 y 的值為何，x 的值均為定值。

這種非左下右上的正相關或非左上右下的負相關形式時，稱為無關或零相關。



焦點 3. 相關係數

從散布圖只能大致看出變數 X 和 Y 之間的相關情況，若想要用比較準確的方式描述變數 X 和 Y 之間的直線相關程度，則可以利用相關係數，以符號 r 表示，其定義如下：

假設 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 代表變數 X, Y 的值，其中 $x_1, x_2, \cdots, x_n$ 不全相等，$y_1, y_2, \cdots, y_n$ 不全相等，而 X 的平均數為 μ_x，Y 的平均數為 μ_y，則 X, Y 的相關係數為 $r=$

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

$$=\frac{x_1'y_1'+x_2'y_2'+\cdots+x_n'y_n'}{n}$$

$x_i'=\dfrac{x_i-\mu_x}{\sigma_x}, y_i'=\dfrac{y_i-\mu_y}{\sigma_y}\ (i=1, 2, \cdots, n)$，分別是 x_i, y_i 的標準分數

註：若令

$$S_{xy}=(x_1-\mu_x)(y_1-\mu_y)+(x_2-\mu_x)(y_2-\mu_y)+\cdots+(x_n-\mu_x)(y_n-\mu_y)$$

$$S_{xx}=(x_1-\mu_x)^2+(x_2-\mu_x)^2+\cdots+(x_n-\mu_x)^2$$

$$S_{yy}=(y_1-\mu_y)^2+(y_2-\mu_y)^2+\cdots+(y_n-\mu_y)^2$$

則 $r=\dfrac{S_{xy}}{\sqrt{S_{xx}}\cdot\sqrt{S_{yy}}}$ 。

焦點 4. 相關係數 r 的性質

(1) $-1\le r\le 1$ 。

(2) $r>0$ 代表正相關，$r<0$ 代表負相關。

(3) r 的值愈接近 1 或 -1，代表直線相關性愈強；
r 的值愈接近 0，代表直線相關性愈弱。

(4) 當 $r=\pm 1$ 時，代表散布圖中的點全都在一條直線上。
若 $r=1$，直線斜率為正，稱為完全正相關；
若 $r=-1$，直線斜率為負，稱為完全負相關。

(5) r 的值所代表的是直線相關的強度，所以 $r=0$ 只代表沒有直線相關或散布圖中的點呈水平線或鉛直線分布，並不代表變數 X 和變數 Y 沒有任何關聯。

焦點 5. 線性變換與相關係數

設兩變數 X, Y 的相關係數為 r，
若變數 X', Y' 滿足 $X'=aX+b$, $Y'=cY+d$ ，且 X', Y' 的相關係數為 r' ，則

(1) 當 $ac>0$ 時，$r'=\dfrac{ac}{|ac|}r=\dfrac{ac}{ac}r=r$

(2) 當 $ac<0$ 時，$r'=\dfrac{ac}{|ac|}r=\dfrac{ac}{-ac}r=-r$

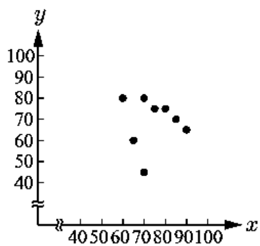
【範例 1】散布圖的描繪

從某班抽樣 8 位同學第一次期中考的數學成績與物理成績的資料如下表：

試畫出這 8 位同學數學與物理成績的散布圖，並判斷兩者是正相關還是負相關。

編號	1	2	3	4	5	6	7	8
數學X（分）	70	65	80	75	60	70	85	90
物理Y（分）	80	60	75	75	80	45	70	65

以 x 坐標表示數學成績，y 坐標表示物理成績，取適當刻度後，可得散布圖如右：

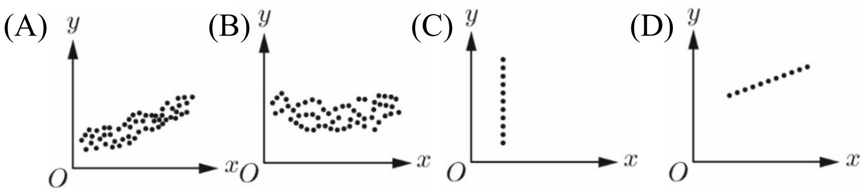


∵散布圖中的點大致呈現從左上方到右下方的趨勢

∴兩者是負相關

【範例 2】相關性的判斷

下列各 X, Y 的散布圖中，試判斷 X 和 Y 的相關程度，選出是正相關的選項。



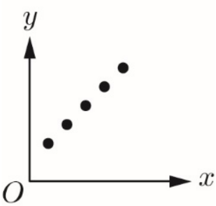
∵當 X, Y 散布圖中的點大致呈現從左下方到右上方的趨勢時，X, Y 為正相關

∴選(A)(D)

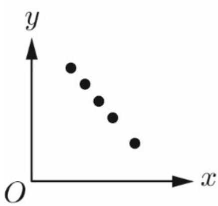
【範例 3】由散布圖判斷相關係數的大小

下列五個散布圖中，x 與 y 的相關係數由小到大排列為何？

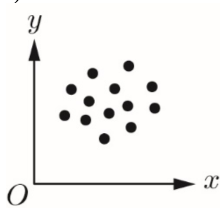
(A)



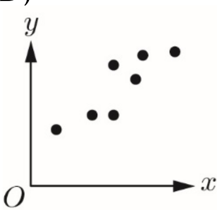
(B)



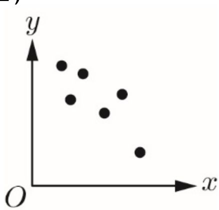
(C)



(D)



(E)



由散布圖可知

(A) $r_A = 1$；

(B) $r_B = -1$；

(C) $r_C \approx 0$；

(D) $0 < r_D < 1$；

(E) $-1 < r_E < 0$

因此，$r_B < r_E < r_C < r_D < r_A$

NOTE

【範例 4】相關係數的計算

有兩個變數 X, Y 的資料如下表：

X	5	6	8	10	12	7	9	7
Y	10	15	15	20	15	5	10	14

試求 X 與 Y 的相關係數。

變數 X 的平均數 $\mu_x = \dfrac{5+6+8+10+12+7+9+7}{8} = 8$

變數 Y 的平均數 $\mu_y = \dfrac{10+15+15+20+15+5+10+14}{8} = 13$

整理後可得下表：

	$x_i-\mu_x$	$y_i-\mu_y$	$(x_i-\mu_x)(y_i-\mu_y)$	$(x_i-\mu_x)^2$	$(y_i-\mu_y)^2$
	−3	−3	9	9	9
	−2	2	−4	4	4
	0	2	0	0	4
	2	7	14	4	49
	4	2	8	16	4
	−1	−8	8	1	64
	1	−3	−3	1	9
	−1	1	−1	1	1
總和	0	0	31	36	144

$\therefore$ 相關係數 $r = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \dfrac{31}{\sqrt{36}\times\sqrt{144}} = \dfrac{31}{6\times 12} = \dfrac{31}{72}$

【範例 5】線性變換與相關係數

已知二變數 X, Y 的 n 筆資料 (x_i, y_i) $(i=1,2,\cdots,n)$，試求：

(1) 若 $y_i = 0.7x_i + 30$ $(i=1,2,\cdots,n)$，則 X 和 Y 的相關係數為何？

(2) 若 $y_i = 3-2x_i$ $(i=1,2,\cdots,n)$，則 X 和 Y 的相關係數為何？

(3) 若 $y_i = 3$ $(i=1,2,\cdots,n)$，則 X 和 Y 的相關係數為何？

(4) 若變數 X 與 Y 的相關係數為 0.7，則：

① 變數 $3X+4$ 與 $2+5Y$ 的相關係數為何？

② 變數 $3X+4$ 與 $-108Y$ 的相關係數為何？

(1) $\because y_i = 0.7x_i + 30$ $(i=1,2,\cdots,n)$

$\therefore (x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，$n$ 個點完全落在一直線

$y = 0.7x + 30$ 上，其斜率為 $0.7 > 0$

即 X, Y 兩變數為完全正相關，故相關係數為 1

(2) $\because y_i = -2x_i + 3$ $(i=1,2,\cdots,n)$

$\therefore (x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，$n$ 個點完全落在一直線

$y = -2x + 3$ 上，其斜率為 $-2 < 0$

即 X, Y 兩變數為完全負相關，故相關係數為 -1

(3) $\because y_i = 3$ $(i=1,2,\cdots,n)$

$\therefore (x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，$n$ 個點完全落在一水平直線

$y = 3$ 上，其斜率為 0

即 X, Y 兩變數為零相關，故相關係數為 0

(4) ① 令 $X' = 3X + 4$ 與 $Y' = 5Y + 2$，X'，Y' 的相關係數 $= r'$

$\because 3 \times 5 > 0$ $\quad \therefore X'$，Y' 的相關係數 $r' = r = 0.7$

② 令 $X' = 3X + 4$ 與 $Y' = -108Y$，X'，Y' 的相關係數 $= r'$

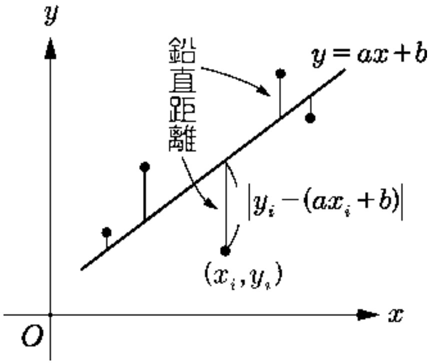
$\because 3 \times (-108) < 0$ $\quad \therefore X'$，Y' 的相關係數 $r' = -r = -0.7$

【主題 2】最小平方法與迴歸直線

焦點 1. 最小平方法

(1) 給定二變數 X, Y 的 n 筆數據 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$。
若 X, Y 的相關係數 r 相當接近 1 或 -1，則代表 X, Y 之間有相當強的直線關聯。

(2) 給定二變數 X, Y 的 n 筆數據 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$。
若直線 $L: y = ax + b$，則點 (x_i, y_i) 與 L 的「鉛直距離」為
$|y_i - (ax_i + b)|$ $(i = 1, 2, \cdots, n)$，如下圖所示。



(3) 當直線 $L: y = ax + b$ 滿足
$[y_1 - (ax_1 + b)]^2 + [y_2 - (ax_2 + b)]^2 + \cdots + [y_n - (ax_n + b)]^2$ 有最小值時，我們稱直線 L 為 Y 對 X 的最小平方迴歸直線，通常簡稱為 Y 對 X 的迴歸直線，亦稱最適直線。

焦點 2. 迴歸直線公式

若變數 X, Y 有 n 筆數據 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，$X$ 的平均數為 μ_x，標準差為 σ_x；Y 的平均數為 μ_y，標準差為 σ_y。X, Y 的相關係數為 r，則 Y 對 X 的迴歸直線公式為

$y - \mu_y = m(x - \mu_x)$ 其中

$$m = r \cdot \frac{\sigma_y}{\sigma_x}$$

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

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

註 1：從上述的公式中可以看出，Y 對 X 的迴歸直線必通過點 (μ_x, μ_y)，且其斜率 m 與相關係數 r 同號。

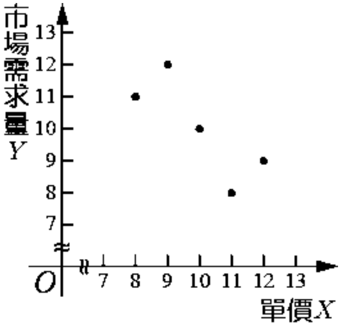
註 2：如果想要利用已知的 x 值來推估或預測 y 的值，我們通常會採取 Y 對 X 的迴歸直線來推估或預測；而如果想要利用已知的 y 值來推估或預測 x 的值，則會採取 X 對 Y 的迴歸直線來推估或預測。

【範例 1】迴歸直線

某肥皂廠商想要推出一種新產品，在產品上市之前先以不同的單價 X（單位：十元）調查市場的需求量 Y（單位：萬盒），調查結果如下：

X	8	9	10	11	12
Y	11	12	10	8	9

(1) 試畫出 X, Y 散布圖。
(2) 試求 Y 對 X 的迴歸直線方程式。
(3) 若將單價定為 105 元，則市場的需求量預估會有幾萬盒？

(1) 

(2) X 平均 $\mu_x=\dfrac{8+9+10+11+12}{5}=10$（十元）

Y 平均 $\mu_y=\dfrac{11+12+10+8+9}{5}=10$（萬盒）

整理後可得下表：

	$x_i-\mu_x$	$y_i-\mu_y$	$(x_i-\mu_x)(y_i-\mu_y)$	$(x_i-\mu_x)^2$
	-2	1	-2	4
	-1	2	-2	1
	0	0	0	0
	1	-2	-2	1
	2	-1	-2	4
總和	0	0	-8	10

$\therefore Y$ 對 X 的迴歸直線方程式為

$$y-10=\frac{-8}{10}(x-10)\Rightarrow y=-\frac{4}{5}x+18$$

(3) 將 $x=10.5$ 代入 Y 對 X 的迴歸直線方程式，得

$$y=-\frac{4}{5}\times10.5+18=9.6$$

即市場的需求量預估會有 9.6 萬盒

【範例 2】迴歸直線

下表是 5 位同學學測中的數學科成績 X（級分）與自然科成績 Y（級分）：

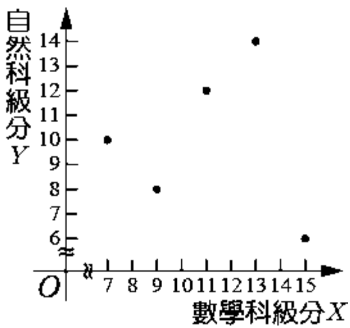
考生	甲	乙	丙	丁	戊
數學科級分X	13	11	9	7	15
自然科級分Y	14	12	8	10	6

(1) 試畫出 X, Y 散布圖。

(2) 試求 Y 對 X 的迴歸直線方程式。

(3) 若已知某生的數學科成績為 12 級分，試預估其自然科成績為多少級分？（四捨五入至整數位）

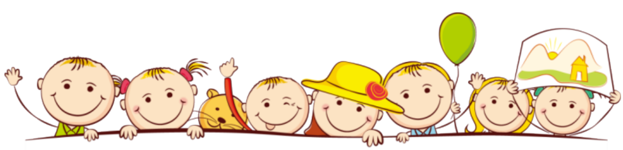
(1)



(2) X 平均 $\mu_x=\dfrac{13+11+9+7+15}{5}=11$（級分）

Y 平均 $\mu_y=\dfrac{14+12+8+10+6}{5}=10$（級分）

整理後可得下表：



	$x_i-\mu_x$	$y_i-\mu_y$	$(x_i-\mu_x)(y_i-\mu_y)$	$(x_i-\mu_x)^2$
	2	4	8	4
	0	2	0	0
	–2	–2	4	4
	–4	0	0	16
	4	–4	–16	16
總和	0	0	–4	40

$\therefore Y$ 對 X 的迴歸直線方程式為 $y-10=\dfrac{-4}{40}(x-11)\Rightarrow$

$$y=-\frac{1}{10}x+\frac{111}{10}$$

(3) 將 $x=12$ 代入 Y 對 X 的迴歸直線方程式，

得 $y=-\dfrac{1}{10}\times 12+\dfrac{111}{10}=9.9\approx 10$

即自然科成績預估為 10 級分

【範例 3】迴歸直線

某校高三第二次期中考數學成績 X 的平均為 60 分，標準差 10 分；英文成績 Y 的平均為 75 分，標準差 15 分。若這兩個科目成績的相關係數為 0.8，則：

(1) 試求英文成績 Y 對數學成績 X 的迴歸直線方程式。

(2) 已知高三學生中數學成績最高為 80 分，試推估其英文成績為幾分？

(1) Y 對 X 的迴歸直線方程式為 $y-75=0.8\times\dfrac{15}{10}\times(x-60)$

$$\Rightarrow y=\frac{6}{5}x+3$$

(2) 將 $x=80$ 代入 Y 對 X 的迴歸直線方程式，得 $y=\frac{6}{5}\times 80+3=99$

即英文成績預估為 99 分

【範例 4】 線性變換與迴歸直線

已知二維數據 (x_i, y_i) $(i=1,2,\cdots,n)$ 中，x_i 的平均數為 $\mu_x=5$，標準差為 $\sigma_x=2$；y_i 的平均數為 $\mu_y=8$，標準差為 $\sigma_y=3$。
若 Y 對 X 的迴歸直線通過點 $(15,20)$，則：

(1) 試求 X 與 Y 的相關係數。

(2) 設 $S=2X+1$, $T=\frac{3}{2}Y-2$，試求 T 對 S 的迴歸直線方程式

(1) $\because$ Y 對 X 的迴歸直線過點 $(15,20)$，又此直線必過點 $(\mu_x,\mu_y)=(5,8)$ $\quad\therefore$ Y 對 X 的迴歸直線方程式為

$$y-8=\frac{20-8}{15-5}(x-5)\Rightarrow y=\frac{6}{5}x+2$$

設 X 與 Y 的相關係數為 r

$$\because\ \frac{6}{5}=m=r\cdot\frac{\sigma_y}{\sigma_x}=r\cdot\frac{3}{2}\qquad\therefore\ r=\frac{4}{5}$$

(2) 令變數 S 的平均數為 μ_s，標準差為 σ_s；變數 T 的平均數為 μ_t，標準差為 σ_t，S 與 T 的相關係數為 r'

則 $\mu_s=2\mu_x+1=2\cdot 5+1=11$，

$\sigma_s=2\sigma_x=2\cdot 2=4$，

$\mu_t=\frac{3}{2}\mu_y-2=\frac{3}{2}\cdot 8-2=10$，

$\sigma_t=\frac{3}{2}\sigma_y=\frac{3}{2}\cdot 3=\frac{9}{2}$， $\quad r'=r=\frac{4}{5}$

因此，T 對 S 的迴歸直線方程式為

$$t-10=\frac{4}{5}\cdot\frac{\frac{9}{2}}{4}\cdot(s-11)=\frac{9}{10}(s-11)$$

$$\Rightarrow t=\frac{9}{10}s+\frac{1}{10}$$

【範例 5】 線性變換與迴歸直線

某班同學的物理成績 X 的平均為 $\mu_x=60$ 分，標準差 $\sigma_x=15$ 分；數學成績 Y 的平均為 $\mu_y=75$ 分，標準差 $\sigma_y=10$ 分。如果 X, Y 的相關係數為 $r=0.75$ ，則：

(1) 試求 Y 對 X 的迴歸直線方程式。

(2) 若任課老師分別調整物理與數學成績：將物理成績乘以 1.4 再減掉 8 分；數學成績乘以 1.2 再減掉 10 分。試求調整完後數學成績對物理成績的迴歸直線方程式。

(1) Y 對 X 的迴歸直線方程式為

$$y-75=0.75\times\frac{10}{15}(x-60)\Rightarrow y=\frac{1}{2}x+45$$

(2) 令 $X'=1.4X-8$, $Y'=1.2X-10$ ，則

X' 的平均數 $\mu_x'=1.4\mu_x-8=1.4\cdot 60-8=76$

標準差 $\sigma_x'=1.4\sigma_x=1.4\cdot 15=21$

Y' 的平均數 $\mu_y'=1.2\mu_y-10=1.2\cdot 75-10=80$

標準差 $\sigma_y'=1.2\sigma_y=1.2\cdot 10=12$

X' 與 Y' 的相關係數 $r'=r=0.75$ ，因此，Y' 對 X' 的迴歸直線方程式為 $y'-80=0.75\cdot\frac{12}{21}(x'-76)=\frac{3}{7}(x'-76)$

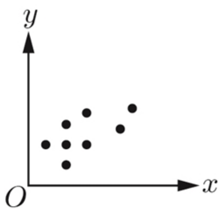
$$\Rightarrow y'=\frac{3}{7}x'+\frac{332}{7}$$

3-2 自我評量

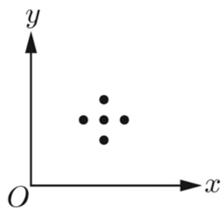
基礎題

1.________下列各散布圖中，試判斷哪些是正相關？

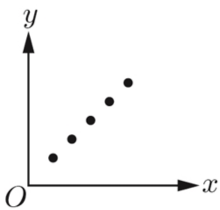
(A)



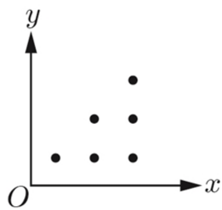
(B)



(C)



(D)

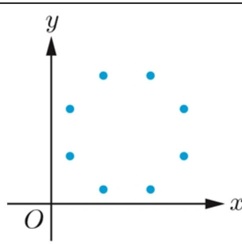


解

(B)沒有直線相關

故選(A)(C)(D)

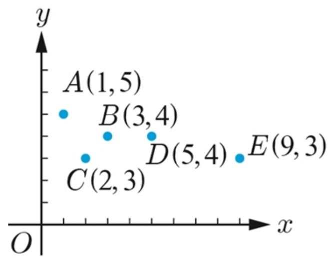
2. ________右圖表兩組數據X,Y的分布圖，試問其相關係數r最接近下列何值？
(A)1　(B)0.5　(C)0　(D)-0.5　(E)-1。



解

散布圖呈現上下左右對稱，其相關係數$r=0$，故選(C)

3.______如右圖所示有 5 筆(X,Y)資料。試問去掉哪一筆資料後，剩下
來
4 筆資料的相關係數最小？
(A) A　(B) B　(C) C　(D) D　(E) E



解

$\mu_x=5, \mu_y=\dfrac{19}{5} \Rightarrow$ 散布圖中，以 $y=\mu_y$ 為新橫軸，$x=\mu_x$ 為新縱軸，則 C 點在第三象限，$(2-\mu_x)(3-\mu_y)>0$，所以去掉 C 點，相關係數最小，故選(C)

4.某班級五位同學的物理成績 X(分)和化學成績 Y(分)如下表所示，則 X 和 Y 的相關係數為__________。(四捨五入至小數點後第二位)

數學成績 X(分)	76	78	78	86	82
化學成績 Y(分)	73	67	64	71	70

解

$$\mu_x=\frac{1}{5}(76+78+78+86+82)=80,$$

$$\mu_y=\frac{1}{5}(73+67+64+71+70)=69$$

	$x_i-\mu_x$	$y_i-\mu_y$	$(x_i-\mu_x)(y_i-\mu_y)$	$(x_i-\mu_x)^2$	$(y_i-\mu_y)$
	–4	4	–16	16	16

	-2	-2	4	4	4
	-2	-5	10	4	25
	6	2	12	36	4
	2	1	2	4	1
總和	0	0	12	64	50

根據相關係數公式得 $\dfrac{12}{\sqrt{64}\times\sqrt{50}}=\dfrac{12}{40\sqrt{2}}=\dfrac{3\sqrt{2}}{20}\approx 0.21$

5.已知 X 與 Y 的相關係數為 -0.65，則 $-2X+3$ 與 $3Y-4$ 這兩組資料的相關係數為________。

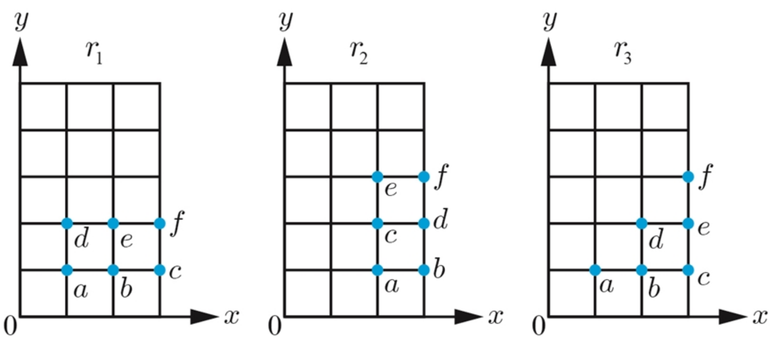
解

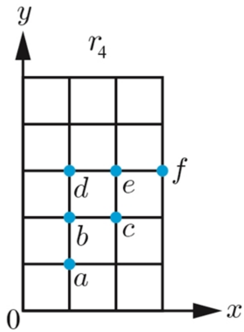
$\because (-2)\times 3<0$

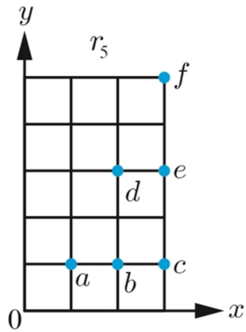
$\therefore -2X+3$ 和 $3Y-4$ 的相關係數 $=-(-0.65)=0.65$

6.________下圖中，有五組數據，每組各有 a,b,c,d,e,f 等六個資料點，設各組的相關係數由左至右分別為 r_1,r_2,r_3,r_4,r_5，則下列關係何者為真？

(A) $r_1=r_2$ (B) $r_2<r_3$ (C) $r_3>r_4$ (D) $r_3<r_5$ (E) $r_4=r_5$







解

①圖1圖形對稱$x=\mu_x$，$y=\mu_y \Rightarrow r_1=0$

②圖2圖形對稱$x=\mu_x$，$y=\mu_y \Rightarrow r_2=0$

③圖3和圖4圖形對稱直線$y=x \Rightarrow r_3=r_4>0$

④圖5兩筆資料X',Y'和圖3的兩筆資料X,Y，滿足$\begin{cases} X'=X \\ Y'=2Y-1 \end{cases}$

$\therefore r_5=r_3=r_4$

故選(A)(B)(E)

7.下表是5位同學學測中的數學與自然科成績：

考生	甲	乙	丙	丁	戊
數學級分X	13	11	9	7	15
自然級分Y	14	12	8	10	6

(1) y對x的迴歸直線方程式為_________。

(2)若某生之數學成績為11級分，試預估自然科級分為____級

解

(1) $\mu_x = \dfrac{1}{5}(13+11+9+7+15) = 11$,

$\mu_y = \dfrac{1}{5}(14+12+8+10+6) = 10$

	$x_i - \mu_x$	$y_i - \mu_y$	$(x_i - \mu_x)(y_i - \mu_y)$	$(x_i - \mu_x)^2$
	2	4	8	4
	0	2	0	0
	–2	–2	4	4
	–4	0	0	16
	4	–4	–16	16
總和	0	0	–4	40

迴歸直線斜率為 $\dfrac{-4}{40} = -\dfrac{1}{10}$

$\therefore$方程式為 $y-10 = -\dfrac{1}{10}(x-11)$，整理得 $y = -\dfrac{1}{10}x + \dfrac{111}{10}$

(2)預估自然級數為 $-\dfrac{1}{10}\times 11 + \dfrac{111}{10} = 10$ 級分

8.設兩組數據 $X: x_1, x_2, \cdots, x_{10}$ 與 $Y: y_1, y_2, \cdots, y_{10}$ 的標準差分別為 $\sigma_X = 1,\ \ \sigma_Y = 5$，若 Y 對 X 的迴歸直線方程式為 $y = 3x+5$，則 X 與 Y 的相關係數為__________。

解

設 X, Y 的相關係數為 r

迴歸直線斜率 $m = 3$

$\because m = r \times \dfrac{\sigma_y}{\sigma_x} \Rightarrow 3 = r \times \dfrac{5}{1} \qquad \therefore r = \dfrac{3}{5} = 0.6$

9.設3筆資料$(x_i,y_i):(6,4),(2,a),(4,3)$，若$Y$對$X$的迴歸直線為$y=0.5x+1$，則實數$a=$________。

解

$\mu_x=\frac{1}{3}(6+2+4)=4$，$\mu_y=\frac{1}{3}(4+a+3)=\frac{7+a}{3}$

$\because$迴歸直線過(μ_x,μ_y) $\therefore\frac{7+a}{3}=0.5\times4+1$，解得$a=2$

10.已知二變數X,Y，其平均數分別為$\mu_x=32$，$\mu_y=56$，且Y對X之迴歸直線L通過點$(57,71)$，則迴歸直線L的斜率為__________。

解

迴歸直線L會通過點$(\mu_x,\mu_y)=(32,56)$，所以迴歸直線L的斜率為$\frac{71-56}{57-32}=\frac{15}{25}=\frac{3}{5}$

進階題

11.設有兩組數據X,Y如右表，已知X的算術平均數為3且X與Y的相關係數為0.5，則$(a,b)=$________。

X	2	a	b	3
Y	3	5	4	4

解

$2+a+b+3=3\times 4 \Rightarrow a+b=7 \Rightarrow b=7-a$

$\mu_x=3, \mu_y=\frac{1}{4}(3+5+4+4)=4$ ，則

	$x_i-\mu_x$	$y_i-\mu_y$	$(x_i-\mu_x)(y_i-\mu_y)$	$(x_i-\mu_x)^2$	$(y_i-\mu_y)^2$
	-1	-1	1	1	1
	$a-3$	1	$a-3$	$(a-3)^2$	1
	$4-a$	0	0	$(4-a)^2$	0
	0	0	0	0	0
總和			$a-2$	$1+(a-3)^2+(4-a)^2$	2

相關係數 $r=\dfrac{a-2}{\sqrt{1+(a-3)^2+(4-a)^2}\cdot\sqrt{2}}=\dfrac{1}{2}$

$\Rightarrow 2(a-2)=\sqrt{2}\cdot\sqrt{2a^2-14a+26}$

$\Rightarrow 4\cdot(a^2-4a+4)=2\cdot(2a^2-14a+26) \Rightarrow 12a=36 \Rightarrow a=3$,

而 $b=7-a=7-3=4$ ，故 $(a,b)=(3,4)$

---

12. 已知二變數 X,Y ，其平均數分別為 $\mu_x=60$， $\mu_y=75$ ，標準差分別為 $\sigma_x=18$， $\sigma_y=15$ ，若 Y 對 X 之迴歸直線過點 $(68,71)$ ，試求
(1)迴歸直線方程式為____________。
(2) X 與 Y 的相關係數 $r=$____________。

解

(1)迴歸直線過點$(60,75)$，$(68,71)$兩點

其斜率$m=\dfrac{71-75}{68-60}=-\dfrac{1}{2}$

$\therefore$方程式為$y-75=-\dfrac{1}{2}(x-60)$，

整理得$y=-\dfrac{1}{2}x+105$

(2)$\because m=r\cdot\dfrac{\sigma_y}{\sigma_x}\Rightarrow -\dfrac{1}{2}=r\times\dfrac{15}{18}$ $\therefore r=-0.6$

13.______某公司為了瞭解所屬員工的健康情形，調查他們的年齡(x)與血壓(y)的數據，經過計算得到血壓(y)對年齡(x)的最適合直線為$y=103+\dfrac{3}{5}x$，年齡(x)與血壓(y)相關係數為0.4，若員工的平均年齡為45歲，年齡標準差為10，則下列敘述何者正確？

(A)員工的平均血壓為130

(B)員工的血壓標準差為6

(C)年齡(x)對血壓(y)的最適合直線必經過點$(45,130)$

(D)某員工的年齡是55歲，則預估其血壓為136

解

(A)○：(μ_x,μ_y)為$y=103+\dfrac{3}{5}x$上一點$\Rightarrow \mu_y=103+\dfrac{3}{5}\times 45=130$

(B)×：$\dfrac{3}{5}=0.4\times\dfrac{\sigma_y}{10}\Rightarrow \sigma_y=15$

(C)○：$(\mu_x,\mu_y)=(45,130)$為迴歸直線上一點

(D)○：預估血壓 $y=103+\frac{3}{5}\times 55=136$　　故選(A)(C)(D)