

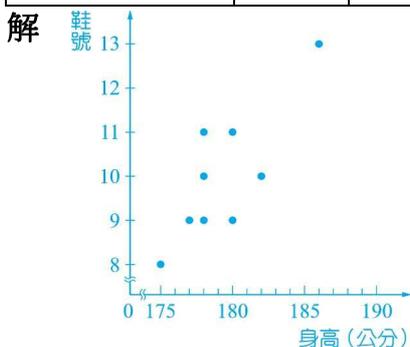
4-2 二維數據分析

重點一 散佈圖

例題 1

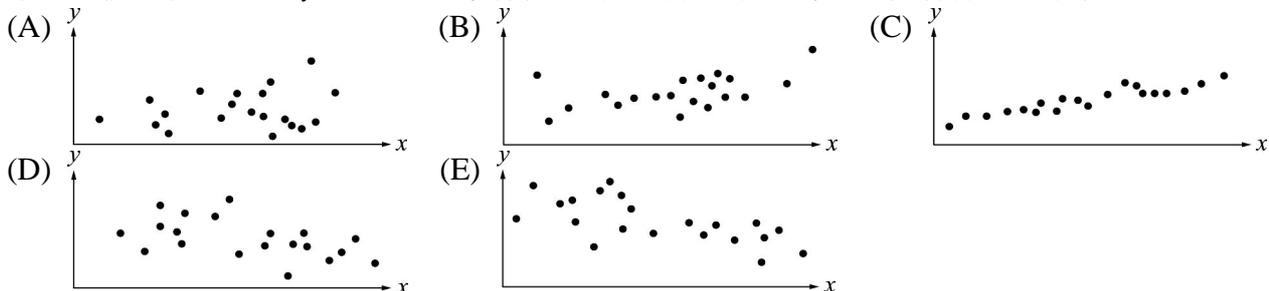
為了解成人的身高與鞋子尺寸的關聯性而做抽樣調查，得到 9 位男性身高與鞋子的號數如下，試作身高與鞋號的散佈圖。

身高 (公分)	186	180	180	182	178	178	175	177	178
鞋 號	13	11	9	10	11	10	8	9	9



例題 2

下列有關兩變量 x 與 y 的 20 對點資料的五個散佈圖中，哪些相關係數為正相關？



解 (A)為正相關
 (B)為正相關
 (C)為正相關
 (D)為負相關
 (E)為負相關
 故選(A)(B)(C)

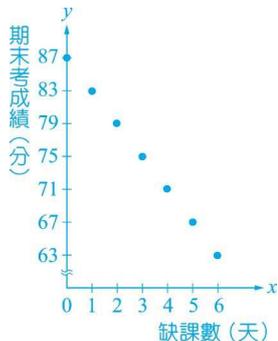
例題 3

某班上有七名學生（以甲、乙、丙、丁、戊、己、庚編號），其期末考成績與該學期上課時缺課數的資料如下：

編號	甲	乙	丙	丁	戊	己	庚
缺課數 x (天)	0	1	2	3	4	5	6
期末考成績 y (分)	87	83	79	75	71	67	63

- (1) 試繪出散佈圖。
- (2) 試繪出標準化數據的散佈圖。

解 (1)



$$(2) \mu_x = \frac{1}{7} (0+1+2+3+4+5+6) = 3$$

$$\mu_y = \frac{1}{7} (87+83+79+75+71+67+63) = 75$$

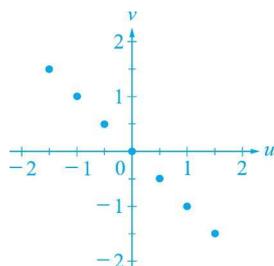
$$\sigma_x = \sqrt{\frac{1}{7} \sum_{i=1}^7 (x_i - \mu_x)^2} = \sqrt{\frac{1}{7} \times [(-3)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2]} = 2$$

$$\sigma_y = \sqrt{\frac{1}{7} \sum_{i=1}^7 (y_i - \mu_y)^2} = \sqrt{\frac{1}{7} [12^2 + 8^2 + 4^2 + 0^2 + (-4)^2 + (-8)^2 + (-12)^2]} = 8$$

標準化後的數據如下：

$u_i = \frac{x_i - \mu_x}{\sigma_x}$	-1.5	-1	-0.5	0	0.5	1	1.5
$v_i = \frac{y_i - \mu_y}{\sigma_y}$	1.5	1	0.5	0	-0.5	-1	-1.5

故得標準化數據的散佈圖如下



重點二 相關係數

例題 4

四人參加性向及成就測驗，其成績如下表所示。

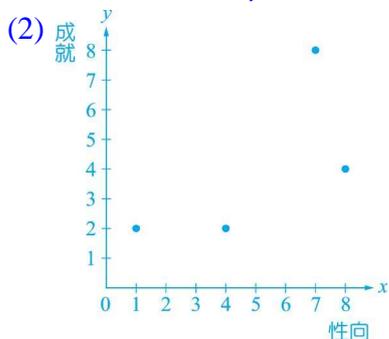
- (1) 若性向、成就測驗成績的算術平均數分別為 μ_x 、 μ_y ，則數對 $(\mu_x, \mu_y) = \underline{\hspace{2cm}}$ 。
- (2) 試作出其散佈圖。
- (3) 性向、成就兩者成績的相關係數為 $\underline{\hspace{2cm}}$ 。(四捨五入取到小數點後第二位)

代 號	A	B	C	D
性向 x	1	8	4	7
成就 y	2	4	2	8

解 (1) $\mu_x = \frac{1+8+4+7}{4} = 5$

$\mu_y = \frac{2+4+2+8}{4} = 4$

故數對 $(\mu_x, \mu_y) = (5, 4)$



(3) $\sum_{i=1}^4 (x_i - \mu_x)^2 = (1-5)^2 + (8-5)^2 + (4-5)^2 + (7-5)^2 = 16 + 9 + 1 + 4 = 30$

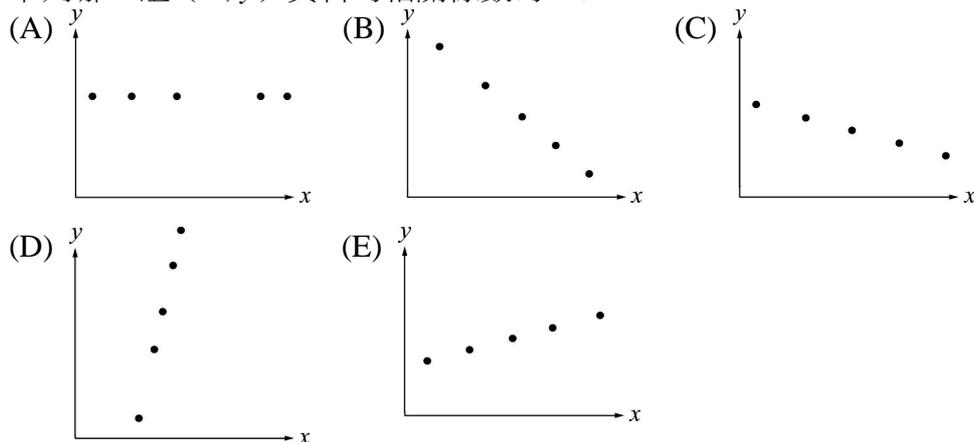
$\sum_{i=1}^4 (y_i - \mu_y)^2 = (2-4)^2 + (4-4)^2 + (2-4)^2 + (8-4)^2 = 4 + 0 + 4 + 16 = 24$

$\sum_{i=1}^4 (x_i - \mu_x)(y_i - \mu_y) = (-4)(-2) + 3 \times 0 + (-1) \times (-2) + 2 \times 4 = 18$

相關係數 $r = \frac{\sum_{i=1}^4 (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^4 (x_i - \mu_x)^2 \cdot \sum_{i=1}^4 (y_i - \mu_y)^2}} = \frac{18}{\sqrt{30 \cdot 24}} \approx 0.67$

例題 5

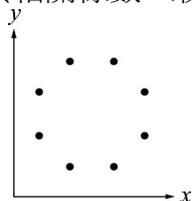
下列哪一組 (x, y) 資料的相關係數為 1？



解 (A)：水平線 \Rightarrow 相關係數為 0
 (B)(C)：直線的斜率為負 \Rightarrow 相關係數為 -1
 (D)(E)：直線的斜率為正 \Rightarrow 相關係數為 1
 故選(D)(E)

例題 6

下圖表兩組數據 x, y 的散佈圖，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 \because 所給八點呈上下對稱、左右對稱
 \therefore 相關係數 $r=0$
 故選(C)

重點三 最小平方與迴歸直線

例題 7

散佈圖上有資料 $(1, 1), (2, 4), (3, 5)$ ，試用最小平方方法求迴歸直線方程式。

解 設所求迴歸直線方程式為 $L: y = a + bx$

要求 a, b 使下圖中的 $d_1^2 + d_2^2 + d_3^2$ 有最小值，計算可得

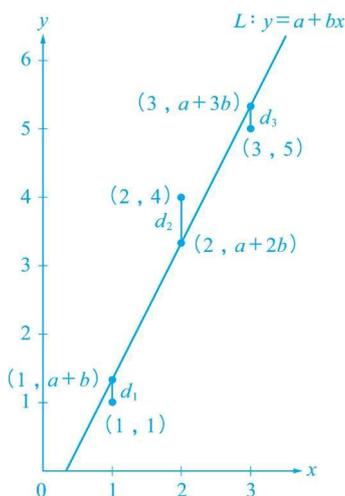
$$\begin{aligned} & d_1^2 + d_2^2 + d_3^2 \\ &= [(a+b) - 1]^2 + [(a+2b) - 4]^2 + [(a+3b) - 5]^2 \\ &= (a+b)^2 - 2(a+b) + 1 + (a+2b)^2 - 8(a+2b) + 16 \\ &\quad + (a+3b)^2 - 10(a+3b) + 25 \\ &= a^2 + 2ab + b^2 - 2a - 2b + 1 + a^2 + 4ab + 4b^2 - 8a - 16b \\ &\quad + 16 + a^2 + 6ab + 9b^2 - 10a - 30b + 25 \\ &= 3a^2 + 12ab + 14b^2 - 20a - 48b + 42 \\ &= 3(a+2b)^2 + 2b^2 - 20a - 48b + 42 \end{aligned}$$

$$= 3 \left[(a+2b) - \frac{10}{3} \right]^2 + 2b^2 - 8b + \frac{26}{3}$$

$$= 3 \left(a+2b - \frac{10}{3} \right)^2 + 2(b-2)^2 + \frac{2}{3}$$

故當 $\begin{cases} a+2b - \frac{10}{3} = 0 \\ b-2 = 0 \end{cases}$ ，即 $\begin{cases} a = -\frac{2}{3} \\ b = 2 \end{cases}$ 時有最小值 $\frac{2}{3}$

故所求的迴歸直線方程式為 $L: y = -\frac{2}{3} + 2x$

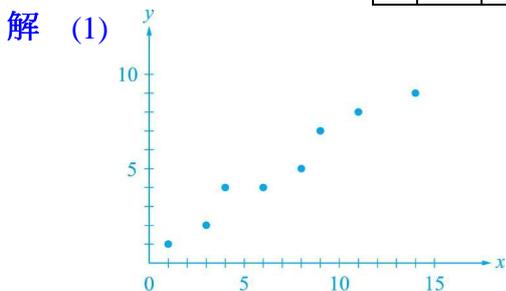


例題 8

設有一群資料如下表，則：

- (1) 試作其散佈圖。
- (2) x 與 y 的相關係數為_____。
(四捨五入取到小數點後第二位)
- (3) y 對 x 的迴歸直線方程式為_____。

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9



(2)

x_i	y_i	$x_i - \mu_x$	$y_i - \mu_y$	$(x_i - \mu_x)(y_i - \mu_y)$	$(x_i - \mu_x)^2$	$(y_i - \mu_y)^2$
1	1	-6	-4	24	36	16
3	2	-4	-3	12	16	9
4	4	-3	-1	3	9	1
6	4	-1	-1	1	1	1
8	5	1	0	0	1	0
9	7	2	2	4	4	4
11	8	4	3	12	16	9
14	9	7	4	28	49	16
合計	56	40		84	132	56

$$\mu_x = \frac{\sum_{i=1}^8 x_i}{8} = \frac{56}{8} = 7, \quad \mu_y = \frac{\sum_{i=1}^8 y_i}{8} = \frac{40}{8} = 5$$

$$\therefore x \text{ 與 } y \text{ 的相關係數 } r = \frac{84}{\sqrt{132 \times 56}} \approx 0.98$$

- (3) 令 y 對 x 的迴歸直線方程式為 $y = a + bx$
由(2)知

$$b = \frac{84}{132} = \frac{7}{11}$$

$$a = \mu_y - b \times \mu_x = 5 - \frac{7}{11} \times 7 = \frac{6}{11}$$

$$\therefore y \text{ 對 } x \text{ 的迴歸直線方程式為 } y = \frac{6}{11} + \frac{7}{11}x$$

例題 9

汽車的耗油量與行駛速度有關聯性，今測試一部 1500 c.c.的汽車，得到下列的數據，

速度 x (公里 / 小時)	60	70	80	90	100
耗油量 y (公里 / 公升)	14	16	19	20	21

試求：

- (1) 耗油量對速度的迴歸直線方程式為_____。
- (2) 預測速度為 85 (公里 / 小時) 時的耗油量為_____公里 / 公升。

解 (1) $\mu_x = \frac{1}{5} (60 + 70 + 80 + 90 + 100) = 80$

$$\mu_y = \frac{1}{5} (14 + 16 + 19 + 20 + 21) = 18$$

x_i	y_i	$x_i - \mu_x$	$y_i - \mu_y$	$(x_i - \mu_x)(y_i - \mu_y)$	$(x_i - \mu_x)^2$
60	14	-20	-4	80	400
70	16	-10	-2	20	100
80	19	0	1	0	0
90	20	10	2	20	100
100	21	20	3	60	400
合計	400	90		180	1000

設迴歸直線方程式為 $y = a + bx$

$$\text{則 } b = \frac{\sum_{i=1}^5 (x_i - \mu_x)(y_i - \mu_y)}{\sum_{i=1}^5 (x_i - \mu_x)^2} = \frac{180}{1000} = \frac{9}{50}$$

$$a = \mu_y - b\mu_x = 18 - \frac{9}{50} \times 80 = \frac{18}{5}$$

故迴歸直線方程式為 $y = \frac{18}{5} + \frac{9}{50}x$

(2) 當 $x = 85$ 時， $y = \frac{18}{5} + \frac{9}{50} \times 85 = 18.9$ (公里 / 公升)