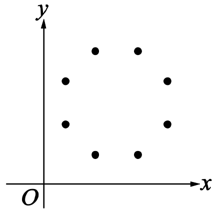
1-4 分析二維數據

例題 1

如右圖表兩組數據 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。

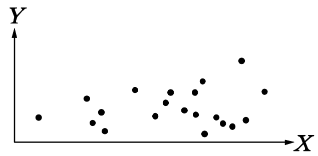


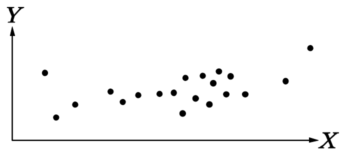
解：∵所給八點呈上下對稱、左右對稱

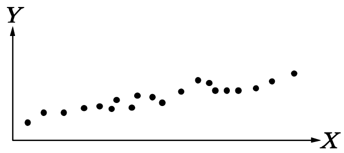
∴相關係數 $r=0$，故選(C)

例題 2

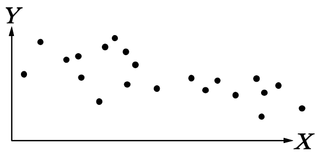
下列有關兩變量 X 與 Y 的 20 對點資料的五個散佈圖中，哪些相關係數為正相關？

(A)


(B)


(C)


(D)


(E)


解：(A)爲低度正相關；(B)爲中度正相關；(C)爲高度正相關；(D)和(E)皆爲中度負相關

故選(A)(B)(C)

例題 3

十位考生之國文與數學成績列表如下：

考生編號	1	2	3	4	5	6	7	8	9	10
國　　文	89	65	76	69	82	57	66	72	78	66
數　　學	75	57	65	65	83	63	58	62	63	69

今算出國文成績之標準差為 9.4 分（取至小數點第一位），數學成績之標準差為 7.9 分（取至小數點第一位），則此十位考生兩科成績之相關係數最接近 (A) -0.85 (B) 0.25 (C) 0.66 (D) 0.78 (E) 0.85。

解：國文平均 $\overline{x}=\dfrac{1}{10}(89+65+76+69+82+57+66+72+78+66)=72$

數學平均 $\overline{y}=\dfrac{1}{10}(75+57+65+65+83+63+58+62+63+69)=66$

又 $\sum\limits_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})=442$

故相關係數 $r=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{(n-1)\times S_x\times S_y}=\dfrac{442}{9\times9.4\times7.9}=0.66\cdots\cdots$，故選(C)

例題 4

四人參加性向及成就測驗，其成績如右表所示。

代　號	A	B	C	D
性向 x	1	8	4	7
成就 y	2	4	2	8

(1) 若性向、成就測驗成績的算術平均數分別為 $\overline{x}$，$\overline{y}$，則數對 $(\overline{x}, \overline{y})=$________。

(2) 性向、成就兩者成績的相關係數為________。

解：(1) $\overline{x}=\dfrac{1+8+4+7}{4}=5$，$\overline{y}=\dfrac{2+4+2+8}{4}=4$

故數對 $(\overline{x}, \overline{y})=(5, 4)$

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

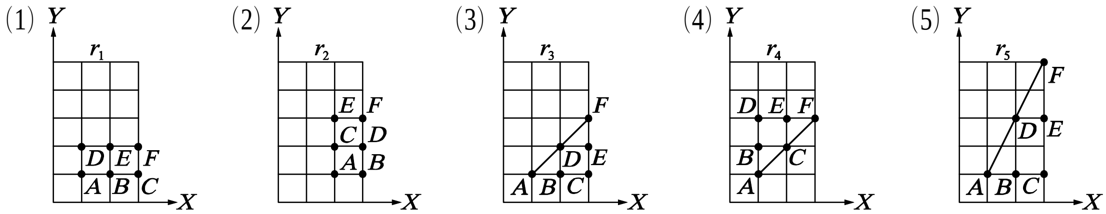
$\sum_{i=1}^{4}(y_i-\overline{y})^2=(2-4)^2+(4-4)^2+(2-4)^2+(8-4)^2=4+0+4+16=24$

$\sum_{i=1}^{4}(x_i-\overline{x})(y_i-\overline{y})=(-4)(-2)+3\times0+(-1)\times(-2)+2\times4=18$

相關係數 $r=\dfrac{18}{\sqrt{30}\cdot\sqrt{24}}\approx0.67$

例題 5

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，則下列關係式何者為真？

(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：圖(1)、(2)之各組資料對稱於點 $(\overline{X}, \overline{Y})$　$\therefore r_1=r_2=0$

圖(3)、(4)之結構相同，且是正相關　$\therefore r_3=r_4>0$

又比較圖(3)、(5)知 $r_{(X,Y)}=r_{(X,2Y-1)}$　$\therefore r_3=r_5$，故選(A)(B)(E)

例題 6

調查全校三年級學生的第一次段考數學成績（x）與物理成績（y），抽查其中 40 位同學，得到平均數、標準差及相關係數如下：$\overline{x}=70$，$\overline{y}=65$，$S_x=10$，$S_y=5$，$r=0.8$。

(1) 求物理成績（y）對數學成績（x）的最佳直線為________。

(2) 已知三年級學生中數學成績最高分為 95 分，試推估其物理成績約為________分。

解：

(1) 設 L：$y=a+bx$，則 $b=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{(x_i-\overline{x})^2}=r\cdot\dfrac{S_y}{S_x}=0.8\cdot\dfrac{5}{10}=0.4$

⇨ $a=\overline{y}-b\overline{x}=65-0.4\times70=37$，故最佳直線 L 爲 $y=37+0.4x$

(2) $\because x=95$　$\therefore y=37+0.4\times95=75$，故物理成績約爲 75 分

例題 7

設有一群資料：

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

，則：

(1) x 對 y 的相關係數為________。

(2) y 對 x 的最佳直線（迴歸線）方程式為________。

解：

	x_i	y_i	$x_i-\overline{x}$	$y_i-\overline{y}$	$(x_i-\overline{x})(y_i-\overline{y})$	$(x_i-\overline{x})^2$	$(y_i-\overline{y})^2$
	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

$\overline{x}=\dfrac{\Sigma x_i}{8}=\dfrac{56}{8}=7$，$\overline{y}=\dfrac{\Sigma y_i}{8}=\dfrac{40}{8}=5$

(1) x 對 y 的相關係數 $r=\dfrac{84}{\sqrt{132}\sqrt{56}}\approx0.98$

(2) $b=\dfrac{84}{132}=\dfrac{7}{11}$

$a=\overline{y}-b\times\overline{x}=5-\dfrac{7}{11}\times7=\dfrac{6}{11}$

$\therefore y$ 對 x 的迴歸線方程式爲 $y=\dfrac{6}{11}+\dfrac{7}{11}x$

例題 8

設有兩組變量 x 與 y，x：x_1，x_2，……，x_{10}；y：y_1，y_2，……，y_{10}，已知

$\sum_{i=1}^{10} x_i=11$，$\sum_{i=1}^{10} x_i^2=32.1$，$\sum_{i=1}^{10} y_i=14$，$\sum_{i=1}^{10} y_i^2=39.6$，$\sum_{i=1}^{10} x_i y_i=28$，則：

(1) 這兩組變量 x 與 y 的相關係數為________。

(2) 滿足這十個點 (x_1,y_1)，(x_2,y_2)，……，(x_{10},y_{10}) 的最適合直線方程式為________。

解：由題意知 $\bar{x}=\frac{11}{10}=1.1$，$\bar{y}=\frac{14}{10}=1.4$

$$\sum_{i=1}^{10}(x_i-\bar{x})^2=\sum_{i=1}^{10}x_i^2-10(\bar{x})^2=32.1-10\times1.1^2=20$$

$$\sum_{i=1}^{10}(y_i-\bar{y})^2=\sum_{i=1}^{10}y_i^2-10(\bar{y})^2=39.6-10\times1.4^2=20$$

$$\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{10}x_iy_i-10(\bar{x})(\bar{y})=28-10\times1.1\times1.4=12.6$$

(1) $r=\dfrac{\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{(x_i-\bar{x})^2}\sqrt{(y_i-\bar{y})^2}}=\dfrac{12.6}{\sqrt{20}\sqrt{20}}=0.63$

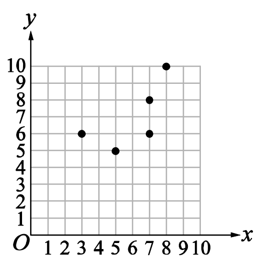
(2) $b=\dfrac{\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})}{(x_i-\bar{x})^2}=\dfrac{12.6}{20}=0.63$

$a=\bar{y}-b\bar{x}=1.4-0.63\times1.1=0.707$

故最適合直線方程式爲 $y=0.63x+0.707$

例題 9

五個學生參加一項包含數學與英文之能力測驗（每科皆為滿分十分）得其成績之散佈圖如右（x 表數學成績，y 表英文成績），試求：



(1) 數學之算術平均數 $\bar{x}$ 為________，標準差 S_x 為________。

(2) x，y 之相關係數為________。

(3) y 對 x 之最佳直線為________。（以 $y=a+bx$ 之形式表示）

解：(1) 由圖知 x，y 之對應關係如下：

x	3	5	7	7	8
y	6	5	6	8	10

可得 $\bar{x}=\frac{1}{5}(3+5+7+7+8)=6$

$\bar{y}=\frac{1}{5}(6+5+6+8+10)=7$

x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\cdot(y_i-\bar{y})$
3	6	-3	-1	9	1	3
5	5	-1	-2	1	4	2
7	6	1	-1	1	1	-1
7	8	1	1	1	1	1
8	10	2	3	4	9	6
				16	16	11

$S_x=\sqrt{\frac{1}{4}(x_i-\bar{x})^2}=\sqrt{\frac{1}{4}(9+1+1+1+4)}=\sqrt{\frac{16}{4}}=2$

(2) 相關係數 $r=\frac{11}{\sqrt{16}\sqrt{16}}=\frac{11}{16}\approx 0.69$

(3) 設最適直線 L：$y=a+bx$，則

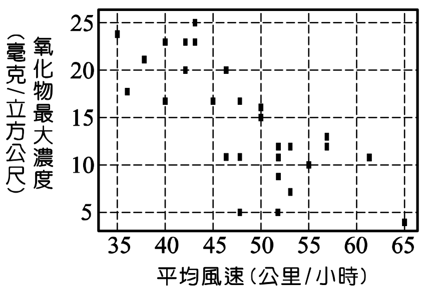
$$b=\frac{\sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y})}{(x_i-\bar{x})^2}=\frac{11}{16}$$

$$a=\bar{y}-b\bar{x}=7-\frac{11}{16}\times 6=\frac{23}{8}$$

$\therefore$最佳直線方程式為 $y=\frac{11}{16}x+\frac{23}{8}$

例題 10

空氣品質會受到汙染物排放量及大氣擴散等因素的影響。某一機構為了解一特定地區的空氣品質，連續二十八天蒐集了該地區早上的平均風速及空氣中某特定氧化物的最大濃度。再繪製這二十八筆資料的散佈圖（見右圖），現根據該圖，可知



(A)此筆資料中，該氧化物最大濃度的標準差大於 15

(B)此筆資料中，該氧化物最大濃度的中位數為 15

(C)此筆資料中，平均風速的中位數介於 45 與 50 間

(D)若以最小平方法決定數據集中直線趨勢的直線，則該直線的斜率小於 0。

解：(A) ╳：最大濃度之平均約爲 15（毫克／立方公尺），而所有的點幾乎集中在 15±10（毫克／立方公尺）內，即 5～25（毫克／立方公尺）

∴標準差不可能大於 15

(B) ╳：最大濃度之樣本由小到大 a_1，a_2，……，a_{28} 共有 28 個

由圖形知最大濃度 $a_{14}\approx 13$，$a_{15}=15$，即中位數 $\dfrac{a_{14}+a_{15}}{2}=14<15$

∴中位數小於 15

(C) ○：平均風速之樣本由小到大 b_1，b_2，……，b_{28} 共有 28 個

由圖形知平均風速之 $b_{13}=b_{14}=b_{15}=48$

即中位數 $\dfrac{b_{14}+b_{15}}{2}=48$　∴中位數介於 45 與 50 之間

(D) ○：根據 28 個點之散佈圖知平均風速與最大濃度兩者爲負相關

∴直線之斜率小於 0

故選(C)(D)