



單元

9

二維數據分析



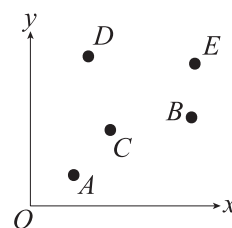
建議配分

每題 10 分。

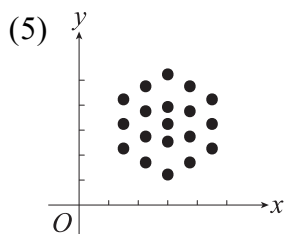
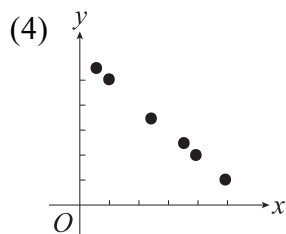
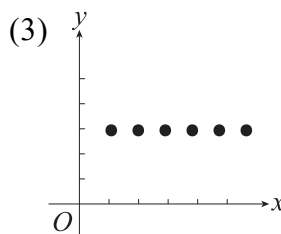
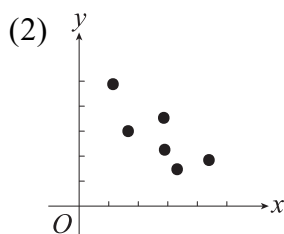
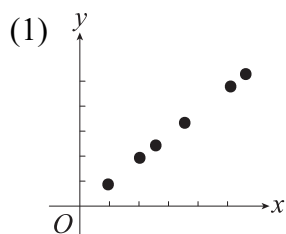
1. 右圖的散布圖中應去掉哪一點後，其相關係數會變大？

(1) A (2) B (3) C (4) D (5) E 。

解► 因為去掉 D 點後，其餘 4 點比較接近一直線，所以相關係數變大。
故選(4)。



2. 下列散布圖中，選出表示兩變量間為負相關的選項。



解► 散布點的分布集中在斜率為負的直線附近即為負相關。
故選(2)(4)。

3. 兩變量 x 與 y 的數據如下表：

x	1	3	5	2	4
y	9	1	9	3	3

求 x 與 y 的相關係數。

解► 兩變量 x 與 y 的平均數分別為

$$\mu_x = \frac{1+3+5+2+4}{5} = 3, \quad \mu_y = \frac{9+1+9+3+3}{5} = 5。$$

依公式需要整理如下表：

$x - \mu_x$	$y - \mu_y$	$(x - \mu_x)^2$	$(y - \mu_y)^2$	$(x - \mu_x)(y - \mu_y)$
-2	4	4	16	-8
0	-4	0	16	0
2	4	4	16	8
-1	-2	1	4	2
1	-2	1	4	-2
總和		$S_{xx} = 10$	$S_{yy} = 56$	$S_{xy} = 0$

代入相關係數的計算公式，得 $r = \frac{S_{xy}}{\sqrt{S_{xx}} \times \sqrt{S_{yy}}} = \frac{0}{\sqrt{10} \times \sqrt{56}} = 0。$

4. 有 20 筆數據 (x_i, y_i) 都滿足 $y_i = -3x_i + 1$ ，其中 $i = 1, 2, \dots, 20$ ，求 x 與 y 的相關係數。

解► 因為數據 (x_i, y_i) 都滿足 $y_i = -3x_i + 1$ ，表示所有點都在一斜直線上，相關係數為 1 或 -1。
又該直線斜率小於 0， x 與 y 為負相關，所以相關係數為 -1。

5. 關於散布圖的敘述，選出正確的選項。

- (1) 若各數據點全落在一斜直線上，表示兩變量呈現完全正相關
- (2) 若以 (μ_x, μ_y) 當作原點，各數據點多半集中在第一、三象限，表示兩變量呈現負相關
- (3) 若各數據點散布上、下、左、右均成對稱，表示兩變量為零相關
- (4) 若各數據點散布在一平行 x 軸的直線上，表示兩變量呈現完全正相關
- (5) 散布圖上各數據點的迴歸直線，其斜率恰等於相關係數。

解►

- (1) 錯誤：可能為完全負相關。
 - (2) 錯誤：兩變量大多呈現正相關。
 - (3) 正確。
 - (4) 錯誤：若平行 x 軸，表示兩變量為零相關。
 - (5) 錯誤：當兩變量的標準差相等時方有此性質。
- 故選(3)。

6. 關於下列的敘述，選出正確的選項。

- (1) 相關係數 r 的變動範圍在 -1 與 1 之間
- (2) 當兩變量 x 與 y 的相關係數愈大時，代表 x 與 y 的相關程度愈強
- (3) 迴歸直線 $y = mx + k$ 中的斜率 m 的變動範圍在 -1 與 1 之間
- (4) 由兩變量 x 與 y 的數據所求得的迴歸直線之斜率 m 與相關係數 r 正負號相同
- (5) 若兩變量 x 與 y 的數據皆滿足 $y = ax + b$ ，則 x 與 y 的相關係數 $r = 1$ 。

解►

- (1) 正確。
- (2) 錯誤：例如相關係數 -1 ，表完全負相關。
- (3) 錯誤：因為迴歸直線 $y = mx + k$ 中的斜率 m 與相關係數 r 的關係為 $m = r \times \frac{\sigma_y}{\sigma_x}$ ，

所以 m 的變動範圍未必在 -1 與 1 之間。

- (4) 正確。
- (5) 錯誤： r 可能為 1 、 -1 或 0 。

故選(1)(4)。

7. 某班學生七人（甲、乙、…、庚）學測數學成績 y （級分）與每週研讀數學時數 x （小時），如下表所示：

學生	甲	乙	丙	丁	戊	己	庚
時數 X	8	16	6	10	12	6	12
成績 Y	9	12	7	11	10	9	12

- (1) 求學測數學成績 Y 與每週研讀數學時數 X 兩者的相關係數。
 (2) 求學測數學成績 Y 對每週研讀數學時數 X 的迴歸直線。

解► 兩變量 x 與 y 的平均數分別為

$$\mu_x = \frac{8+16+6+10+12+6+12}{7} = 10, \quad \mu_y = \frac{9+12+7+11+10+9+12}{7} = 10。$$

依公式需要整理如下表：

$x - \mu_x$	$y - \mu_y$	$(x - \mu_x)^2$	$(y - \mu_y)^2$	$(x - \mu_x)(y - \mu_y)$
-2	-1	4	1	2
6	2	36	4	12
-4	-3	16	9	12
0	1	0	1	0
2	0	4	0	0
-4	-1	16	1	4
2	2	4	4	4
總和		$S_{xx} = 80$	$S_{yy} = 20$	$S_{xy} = 34$

- (1) 代入相關係數的計算公式，得

$$r = \frac{S_{xy}}{\sqrt{S_{xx}} \times \sqrt{S_{yy}}} = \frac{34}{\sqrt{80} \times \sqrt{20}} = \frac{17}{20} = 0.85。$$

- (2) 代入迴歸直線方程式 $y - \mu_y = \frac{S_{xy}}{S_{xx}}(x - \mu_x)$ ，得 $y - 10 = \frac{34}{80}(x - 10)$ ，

$$\text{即 } y = \frac{17}{40}x + \frac{23}{4}。$$

8. 設有 n 組數據 (x_i, y_i) ， $i = 1, 2, \dots, n$ ，兩變量 x 與 y 的算術平均數分別為 $\mu_x = 5$ ， $\mu_y = 3$ ，標準差分別為 $\sigma_x = 2$ ， $\sigma_y = 6$ 。已知 y 對 x 的迴歸直線過點 $(2, 6)$ ，求兩變量 x 與 y 的相關係數。

解► 迴歸直線方程式 $y - \mu_y = m(x - \mu_x)$ ，其中 $m = r \times \frac{\sigma_y}{\sigma_x}$ 。

因為 y 對 x 的迴歸直線方程式為 $y - 3 = m(x - 5)$ ，且通過點 $(2, 6)$ ，

所以 $6 - 3 = m(2 - 5)$ ，解得 $m = -1$ 。

又 $-1 = r \times \frac{6}{2}$ ，得 $r = -\frac{1}{3}$ 。故兩變量 x 與 y 的相關係數為 $-\frac{1}{3}$ 。

9. 某班學生的身高 x 的平均數 $\mu_x = 160$ 公分、標準差 $\sigma_x = 10$ 公分；體重 y 平均數 $\mu_y = 50$ 公斤、標準差 $\sigma_y = 8$ 公斤。已知身高和體重的相關係數 $r = 0.7$ 。

(1) 求體重 y 對身高 x 的迴歸直線方程式。

(2) 利用迴歸直線預測：某人的身高為 170 公分，其體重約為多少公斤？

(3) 當將身高單位改為 x' 公吋時，求身高 x' 公吋與體重 y 公斤的相關係數。

解► (1) 迴歸直線方程式 $y - \mu_y = m(x - \mu_x)$ ，其中 $m = r \times \frac{\sigma_y}{\sigma_x} = 0.7 \times \frac{8}{10} = 0.56$ ，

故迴歸直線為 $y - 50 = 0.56(x - 160)$ ，即 $y = 0.56x - 39.6$ 。

(2) 將 $x = 170$ 代入 $y - 50 = 0.56(x - 160)$ ，得 $y = 50 + 5.6 = 55.6$ （公斤）。

(3) 相關係數與單位無關。改變 x 的度量單位時，相關係數不會改變。

故 x' 與 y 的相關係數為 0.7。

10. 某飲料店根據過去的銷售紀錄，當每日最高氣溫在 22°C 到 39°C 時，該日飲料的銷售量與當天的最高氣溫之相關係數為 0.99，部分紀錄如下表：

最高氣溫 $^\circ\text{C}$	25	27	29	31	33	35
銷售量（杯）	205	257	303	356	408	464

已知某日最高氣溫為 38°C ，依據上述的資訊推測，試問該日飲料的銷售量應接近下列哪個選項？

- (1) 490 杯 (2) 520 杯 (3) 542 杯 (4) 616 杯。

解► 若以最高氣溫當橫軸，銷售量當縱軸，作散布圖時，因為相關係數為 0.99，所以散布圖的所有點會相當靠近一條斜率為正的直線 L 。觀察紀錄表中銷售量的變化：

最高氣溫 $^\circ\text{C}$	25	27	29	31	33	35
銷售量（杯）	205	257	303	356	408	464

$\underbrace{\hspace{1.5cm}}_{+52}$
 $\underbrace{\hspace{1.5cm}}_{+46}$
 $\underbrace{\hspace{1.5cm}}_{+53}$
 $\underbrace{\hspace{1.5cm}}_{+52}$
 $\underbrace{\hspace{1.5cm}}_{+56}$

得知：最高氣溫每增加 2°C ，銷售量約增加 52 杯。

因此，直線 L 的斜率約為 $\frac{52}{2} = 26$ 。

令 38°C 時的銷售量為 x 杯。由直線斜率的定義，得 $\frac{x - 464}{38 - 35} \approx 26$ ，

解得 $x \approx 542$ ，故選(3)。