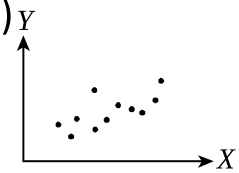
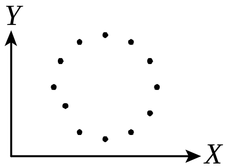
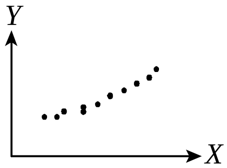
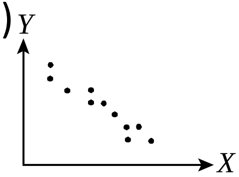
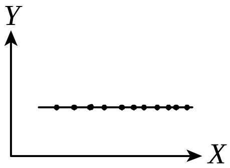
6-2 二維數據分析(常考題型 1)

下列有(A)到(E)五個散布圖，哪些呈現正相關？哪些呈現負相關？哪些呈現零相關？

(A)

(B)

(C)

(D)

(E)


解答　正相關：(A)(C)；負相關：(D)；零相關：(B)(E)

解析　由散布圖判斷，正相關：(A)(C)；負相關：(D)；零相關：(B)(E)．

6-2 二維數據分析(常考題型 2)

某肥皂廠商欲推出一種新產品，在上市前以不同的單價 x（單位：十元）調查市場的需求量 y(單位：萬盒)，調查結果如表，問 x 和 y 的相關係數接近下列哪一個值？

(1)$\frac{4}{5}$　(2)$\frac{2}{5}$　(3)0　(4)$-\frac{2}{5}$　(5)$-\frac{4}{5}$．

x	8	9	10	11	12
y	11	12	10	8	9

解答　5

解析

x_i	y_i	$x_i-\overline{x}$	$y_i-\overline{y}$	$(x_i-\overline{x})(y_i-\overline{y})$	$(x_i-\overline{x})^2$	$(y_i-\overline{y})^2$
8	11	-2	1	-2	4	1
9	12	-1	2	-2	1	4
10	10	0	0	0	0	0
11	8	1	-2	-2	1	4
12	9	2	-1	-2	4	1
50	50			-8	10	10

$\overline{x}=\frac{1}{5}(8+9+10+11+12)=10$，$\overline{y}=\frac{1}{5}(11+12+10+8+9)=10$，

$$r=\frac{\sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{5}(x_i-\overline{x})^2\cdot\sum_{i=1}^{5}(y_i-\overline{y})^2}}=\frac{-8}{\sqrt{10\times10}}=-\frac{8}{10}=-\frac{4}{5}$$，故選(5)．

6-2 二維數據分析(常考題型 3)

(1)證明相關係數的性質：a，b，p，q 均為實數，令 $W=pX+a$，$Z=qY+b$，設 $r_{(x,y)}$表示 X 與 Y 的相關係數，$r_{(w,z)}$表示 W 與 Z 的相關係數，則 ①若 $pq>0$，則 $r_{(w,z)}=r_{(x,y)}$ ． ②若 $pq<0$，則 $r_{(w,z)}=-r_{(x,y)}$ ． (2)設變數 X 與 Y 的相關係數 $r_{(x,y)}=0.56$，若 $X'=1000X+600$，$Y'=-500Y-300$，則變數 X'與 Y'的相關係數 $r_{(x',y')}=$ ？	

解答　(1)見解析;(2)-0.56

解析　(1)由於

$$r_{(w,z)}=\frac{\sum_{i=1}^{n}(w_i-\overline{w})(z_i-\overline{z})}{\sqrt{\sum_{i=1}^{n}(w_i-\overline{w})^2\sum_{i=1}^{n}(z_i-\overline{z})^2}}=\frac{\sum_{i=1}^{n}(px_i+a-(p\overline{x}+a))(qy_i+b-(q\overline{y}+b))}{\sqrt{\sum_{i=1}^{n}(px_i+a-(p\overline{x}+a))^2\sum_{i=1}^{n}(qy_i+b-(q\overline{y}+b))^2}}$$

$$=\frac{\sum_{i=1}^{n}(px_i-p\overline{x})(qy_i-q\overline{y})}{\sqrt{\sum_{i=1}^{n}(px_i-p\overline{x})^2\sum_{i=1}^{n}(qy_i-q\overline{y})^2}}=\frac{pq}{|pq|}\cdot\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}}=\frac{pq}{|pq|}\cdot r_{(x,y)}，$$

①若 $pq>0$，$\frac{pq}{|pq|}=\frac{pq}{pq}=1$，則 $r_{(w,z)}=r_{(x,y)}$ ．

②若 $pq<0$，$\frac{pq}{|pq|}=\frac{pq}{-pq}=-1$，則 $r_{(w,z)}=-r_{(x,y)}$ ．

(2)$pq=1000\times(-500)<0$，故 $r_{(x',y')}=-r_{(x,y)}=-0.56$ ．

6-2 二維數據分析(常考題型 4)

有 20 筆數據(x_i,y_i)，$i=1$，2，…，20．其平均$\bar{x}=3$，$\bar{y}=5$，x 與 y 的相關係數 $r=0.8$，且 y 對 x 的迴歸直線通過點(2,0)．選出正確的選項． (1)X 與 Y 為正相關 (2)迴歸直線通過點(3,5) (3)迴歸直線的斜率為 0.8 (4)迴歸直線通過點(4,10) (5)x 的標準差小於 y 的標準差．	

解答　1245

解析　(1)相關係數為正數，所以 X 與 Y 為正相關

(2)迴歸直線必過$(\bar{x},\bar{y})=(3,5)$．

(3)已知迴歸直線通過$(\bar{x},\bar{y})=(3,5)$及(2,0)，因此直線斜率為$\dfrac{5-0}{3-2}=5$．

(4)由(3)得迴歸直線的方程式為 $y-5=5(x-3)\Rightarrow y=5x-10$．

將 $x=4$ 代入迴歸直線方程式得 $y=5\cdot 4-10=10$，迴歸直線過點(4,10)．

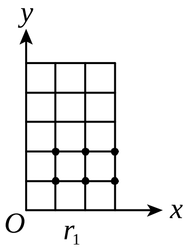
(5)因為迴歸直線的斜率為$r\cdot\dfrac{\sigma_y}{\sigma_x}$，所以$0.8\times\dfrac{\sigma_y}{\sigma_x}=5\Rightarrow\sigma_y=\dfrac{25}{4}\sigma_x$．

因此 x 的標準差小於 y 的標準差．故選項(1)(2)(4)(5)正確．

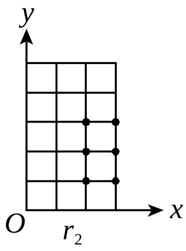
6-2 二維數據分析(常考題型 5)

下圖中有五組數據，每組各有六個資料點．各組的相關係數分別為 r_1，r_2，r_3，r_4，r_5，試比較其大小關係．

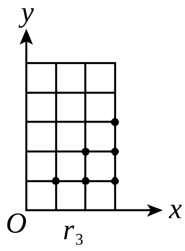
(1)



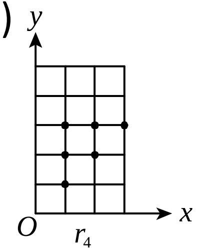
(2)



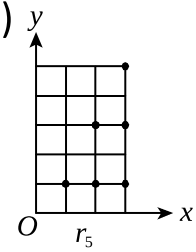
(3)



(4)

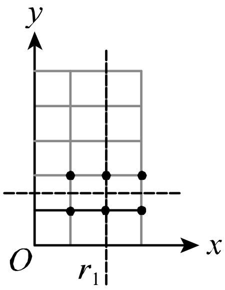


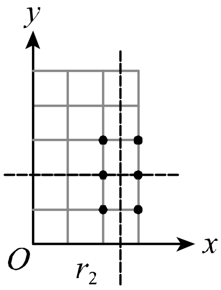
(5)



解答　$r_1 = r_2 < r_3 = r_4 = r_5$

解析　比較 r_1 與 r_2，作 $\bar{x}$ 與 $\bar{y}$ 線如圖所示：





所有點對稱分布在四個象限，其 $(x_i-\mu_x)(y_i-\mu_y)$ 的總和為 0，故 $r_1 = r_2 = 0$．

圖(3)(4)(5)中，當 x 增加時，y 有增加趨勢，其相關係數均大於 0，將圖(3)中的點 x 坐標與 y 坐標互換，恰為圖(4)中的點，故 $r_4 = r_3$，將圖(3)中的點 x 坐標保持不變，而 y 坐標乘以 2 再減去 1，即為圖(5)中的點，故 $r_5 = r_3$．

答案為 $r_1 = r_2 < r_3 = r_4 = r_5$．

6-2 二維數據分析(常考題型 6)

某肥皂廠商欲推出一種新產品,在上市以前以不同的單價 x(單位: 10 元), 調查市場的需求量 y (單位: 萬盒). 調查結果如下:

x	8	9	11	12
y	11	12	8	9

(1)試求 y 對 x 的迴歸直線為____________ .

(2)當單價定為 100 元時,

請預測其銷售量為____________萬盒 .

解答 (1) $y=-\frac{4}{5}x+18$;(2)10 萬盒

解析 (1)（利用公式）因為 $\overline{x}=10$， $\overline{y}=10$， 且

x	8	9	11	12	和
y	11	12	8	9	
$x-\overline{x}$	–2	–1	1	2	
$y-\overline{y}$	1	2	–2	–1	
$(x-\overline{x})^2$	4	1	1	4	10
$(x-\overline{x})(y-\overline{y})$	–2	–2	–2	–2	–8

$m=\dfrac{\sum_{i=1}^{4}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{4}(x_i-\overline{x})^2}=\dfrac{-8}{10}=-\dfrac{4}{5}$， 故 y 對 x 的迴歸直線為

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

(2)當 $x=10$ 時，得 $y=-\frac{4}{5}\times 10+18=10$ （萬盒）.

6-2 二維數據分析(常考題型 7)

某一實驗測得 15 組樣本點(x_1,y_1)，(x_2,y_2)，…，(x_{15},y_{15})，已知 $\sum_{i=1}^{15}x_i=165$，$\sum_{i=1}^{15}y_i=315$，$\sum_{i=1}^{15}(x_i-\bar{x})^2=64$，$\sum_{i=1}^{15}(y_i-\bar{y})^2=144$，及$\sum_{i=1}^{15}(x_i-\bar{x})(y_i-\bar{y})=0$，則

(1)x 與 y 的相關係數為_____________．

(2)Y 對 X 的迴歸直線為_____________．

(3)若再增加一組數據(20,21)，則滿足這 16 個數據之 Y 對 X 的迴歸直線為_____________．

解答　(1)0;(2)$y=21$;(3)$y=21$

解析　(1)$r=\dfrac{\sum_{i=1}^{15}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{15}(x_i-\bar{x})^2\cdot\sum_{i=1}^{15}(y_i-\bar{y})^2}}=0$．

(2)$m=r\cdot\dfrac{\sigma_y}{\sigma_x}=0$，故 y 對 x 的迴歸直線為 $y-21=0\Rightarrow y=21$．

(3)增加的點(20,21)也在直線 $y=21$ 上，並不影響，

故這 16 個數據之 y 對 x 的迴歸直線仍為 $y=21$．

6-2 二維數據分析(常考題型 8)

設(3 , 1)，(5 , 3)，(1 , t)為三筆(x , y)的數據資料，其 Y 對 X 的迴歸直線為 $y=\frac{1}{4}x+\frac{5}{4}$，求 t 之值．

解答　2

解析　兩變量的算術平均數

$\bar{x}=\dfrac{3+5+1}{3}=3$，$\bar{y}=\dfrac{1+3+t}{3}=\dfrac{4+t}{3}$

由於 Y 對 X 的迴歸直線必過此三筆資料的平均數$(\bar{x},\bar{y})$，

代入得$\dfrac{4+t}{3}=\dfrac{1}{4}\times3+\dfrac{5}{4}$，整理得 $t=2$．

6-2 二維數據分析(常考題型 9)

高三某班學生的身高平均 $\mu_x=165$ 公分，身高的標準差 $\sigma_x=10$ 公分；體重平均 $\mu_y=50$ 公斤，標準差 $\sigma_y=8$ 公斤．而身高和體重的相關係數 $r=0.72$． (1)求體重 y 對身高 x 的迴歸直線方程式． (2)若已知某人的身高為 175 公分，我們可預測：其體重約為多少公斤? (3)若將身高單位改為吋，新的身高（x'吋）與體重（y 公斤）的相關係數為多少?	

解答　(1)$y-50=0.576(x-165)$;(2)55.76 公斤;(3)0.72

解析　(1)迴歸直線方程式 $y-\mu_y=m(x-\mu_x)$，其中 $m=r\cdot\dfrac{\sigma_y}{\sigma_x}=0.72\times\dfrac{8}{10}=0.576$，

故迴歸直線為 $y-50=0.576(x-165)$．（或 $y=0.576x-45.04$）．

(2)$x=175$ 代入 $y-50=0.576(x-165)$，得 $y=50+5.76=55.76$（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 $r=0.72$．