7-1 一維數據分析

1. 集中趨勢數⇒ 具代表性

一筆數據資料有 n 個數值 $x_1,x_2,\cdots,x_n$ ，假設 $x_1\le x_2\le\cdots\le x_n$

(1)算術平均數：$\mu=\dfrac{x_1+x_2+\cdots+x_n}{n}$ 。

(2)第 k 百分位數：P_k 表示有 $k\%$ 的數據小於或等於 P_k ，且有 $(100-k)\%$ 的數據大於或等於 P_k 。

【求 P_k】(1)若 $a=n\times k\%$ 是整數，則 $P_k=\dfrac{x_a+x_{a+1}}{2}$ 。

(2)若 $a=n\times k\%$ 不是整數，則 $P_k=x_{[a]+1}$ 。

(3)幾何平均數：$G.M=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}$ 。

(4)眾數：M_0 表示出現次數最多的數值。

(5)中位數：M_e 表示資料由小到大，最中間的數。(第 50 百分位數 P_{50})

(6)四分位數：Q_1,Q_2,Q_3 稱為第一四分位數、第二四分數(中位數)及第三四分位數，分別表示在數據中的 $\frac{1}{4},\frac{2}{4},\frac{3}{4}$ 的數值。其值分別為 $Q_1=P_{25}$ 、$Q_2=P_{50}$ 、$Q_3=P_{75}$ 。

2. 分散趨勢數⇒ 分散程度 (數值恆正，值越大表越 分散)

一筆數據資料有 n 個數值 $x_1,x_2,\cdots,x_n$ ，假設 $x_1\le x_2\le\cdots\le x_n$

(1)全距：$R=x_n-x_1$ 表示最大和最小數值的差。

(2)標準差：$\sigma=\sqrt{\dfrac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{n}}=\sqrt{\dfrac{x_1{}^2+x_2{}^2+\cdots+x_n{}^2}{n}-\mu^2}$ 。

【想法】大概是與平均距離的平均

(3)變異數：σ^2 。

3. 一維數據的伸縮與平移

	集中趨勢數			分散趨勢數	
統計量	平均	眾數	中位數	全距	標準差
X	μ	M_0	M_e	L	σ
$aX+b$	$a\mu+b$	aM_0+b	aM_e+b	$\lvert a\rvert\cdot L$	$\lvert a\rvert\cdot\sigma$

4. 數據標準化⇒①平均為 0 ②標準差為 1

一筆數據資料有 n 個數值 $x_1,x_2,\cdots,x_n$ ，其平均為 μ ，標準差為 σ 。將一筆數據每一個數值，減去其平均，再除以標準差，稱之為標準化。任一數值 x_i 的標準分數 $z_i=\dfrac{x_i-\mu}{\sigma}$ 。

EXAMPLE 1

在某項才藝競賽中，為了避免評審個人主觀影響參賽者成績太大，主辦單位規定：先將 15 位評審給同一位參賽者的比賽成績求得算術平均數，再將與平均數相差超過 15 分的評審成績剔除後重新計算平均值做為此參賽者的比賽成績。現在有一位參賽者所獲 15 位評審的平均成績為 76 分，其中有三位評審給的成績 92、45、55 應剔除，求這個參賽者的比賽成績。

㊣平均 ⇒ 總分

15人總分為 $15\times 76=1140$

剔除3人後 12人總分為 $1140-92-45-55=948$

12人平均 $=\dfrac{948}{12}=79$ #

EXAMPLE 2

小明觀察一組正整數，其中 1 個 1、2 個 2、…、30 個 30，則下列敘述何者正確？

(1)此數據的眾數為 30

(2)此數據的中位數為 20

(3)此數據的算術平均數是正整數

(4)此數據的中位數大於算術平均數

共有 $1+2+\cdots+30=\dfrac{30\times 31}{2}=465$ 個數

這465個數總和為 $=1\times 1+2\times 2+\cdots+30\times 30$

$=\dfrac{30\times 31\times 61}{6}=9455$

(1) 眾數為 30

(2) 中位數為第 233個數

$1+2+\cdots+n\approx 233$,

$1+2+\cdots+21=231$

$1+2+\cdots+21+22=253$

故第233個數為22

(3) 平均 $=\dfrac{9455}{465}=20.\overline{3}$

(4) $22>20.\overline{3}$

選(1)(4) #

EXAMPLE 3

下表為某 30 位同學在某次段考時數學的成績，由小排到大。

57,59,60,68,70,72,75,76,77,78,78,81,82,82,82,83,84,87,88,90,90,90,90,90,93,93,94,96,96,99，試求：

(1)全班的平均分數

(2)若頂標、前標、均標、後標及底標分別定義為第 88 百分位數 P_{88}、第 75 百分位數 P_{75}、第 50 百分位數 P_{50}、第 25 百分位數 P_{25}、第 12 百分位數 P_{12}，則這五標的分數各為多少分？

(3)小惠的分數恰好是平均分數，則她達到哪一標的水準？

(A)頂標 (B)前標 (C)均標 (D)後標 (E)底標

(1) $90+\dfrac{-33-31-30-22-20-18-15-14-13-12-12-9-8-8-8-7-6-3-2+3+3+4+6+6+9}{30}=82.9$

(2) $30\times 0.88=26.4$，$P_{88}=x_{27}=94$

$30\times 0.75=22.5$，$P_{75}=x_{23}=90$

$30\times 0.5=15$，$P_{50}=\dfrac{x_{15}+x_{16}}{2}=\dfrac{82+83}{2}=82.5$

$30\times 0.25=7.5$，$P_{25}=x_8=76$

$30\times 0.12=3.6$，$P_{12}=x_4=68$

(3) $82.5<82.9<90$

故為均標，選(C) #

EXAMPLE 4

小王沉迷於手機遊戲，於是媽媽規定小王必須降低手機的網路流量，平均每個月減少 40%。已知規定的前三個月，網路流量分別較前一個月減少 20%、40%、25%，則第四個月較前一個月減少 x%，小王才能達成規定，求 x 值。

設一開始使用量 N

媽媽規定 實際

$N(1-40\%)^4=N(1-20\%)(1-40\%)(1-25\%)(1-x\%)$

$\Rightarrow (0.6)^3=0.8\cdot 0.75\cdot (1-x\%)$

$\Rightarrow 1-x\%=0.36$，$x\%=64\%$，$x=64$ #

EXAMPLE 5

已知 9 筆數據，經標準化後之值分別為

0.6，0.7，0.8，−0.1，−1.5，−2，1，0，x 。

原來 9 筆數據的算術平均數為 50，標準差為 20，求原始數據的中位數。

㊣標準化 ⇒ 平均0，標準差 1

$0.6+0.7+0.8-0.1-1.5-2+1+0+x=0$

$\Rightarrow x=0.5$

設原始分數 y，$\dfrac{y-50}{20}=0.5$，$y=60$ #

EXAMPLE 6

下列 5 組資料，何者標準差最大？

(1) 1,1,1,1,1,10,10,10,10,10

(2) 1,1,1,1,1,5,5,5,5,5

(3) 4,4,4,5,5,5,5,6,6,6

(4) 1,1,2,2,3,3,4,4,5,5

(5) 1,2,3,4,5,6,7,8,9,10

越分散，標準差越大。

選(1)#

EXAMPLE 7

某生第一次段考考六科的平均為 80 分，若已知其中五科成績為 68,80,80,80,86，則六科成績的標準差為何？

設第六科分數為 x

$\frac{68+80+80+80+86+x}{6}=80$，$x=86$

標準差 $=\sqrt{\frac{(-12)^2+0^2+0^2+0^2+6^2+6^2}{6}}=6$

EXAMPLE 8

某校高一第一次段考數學成績不太理想，多數同學成績偏低。考慮到可能是同學們適應不良所致，數學老師決定將每人的原始成績取平方根後再乘以 10 作為正式紀錄的成績。今隨機抽選 100 位同學，發現調整後的成績其平均為 65 分，標準差為 15 分。試問這 100 位同學未調整前的成績之平均 M 介於哪兩個連續正整數之間。

設原始分數 x_i ⇒ 新分數 $10\sqrt{x_i}$

新分數標準差 $=\sqrt{\frac{(10\sqrt{x_1})^2+(10\sqrt{x_2})^2+\cdots+(10\sqrt{x_{100}})^2}{100}-65^2}=15$

⇒ $x_1+x_2+\cdots+x_{100}-65^2=15^2$

⇒ $x_1+x_2+\cdots+x_{100}=4225+225=4450$

⇒ $M=\frac{x_1+x_2+\cdots+x_{100}}{100}=44.5$

$44<M<45$ #

EXAMPLE 9

陳老師將本學期任教甲、乙兩班的測驗成績結果如表所示。陳老師想將兩班共 50 人的成績合併，下列敘述何者正確？

(1)合併之算術平均數高於 70 分　(2)合併之算術平均數介於 60 至 70 分之間

(3)合併之中位數低於 61 分　(4)合併之中位數介於 61 至 68 分之間

(5)合併之標準差低於 8 分

【想法】平均 $\mu \Rightarrow$ 總分 $x_1+x_2+\cdots+x_n$

標準差 $\sigma \Rightarrow$ 平方和 $x_1^2+x_2^2+\cdots+x_n^2$

	甲	乙
人數	30	20
平均	70	60
中位數	68	61
標準差	8	10

(1) 甲班：$x_1, x_2, \cdots, x_{30}$

(2) 乙班：$x_{31}, x_{32}, \cdots, x_{50}$

$x_1+x_2+\cdots+x_{30}=30\times70=2100$

$x_{31}+x_{32}+\cdots+x_{50}=20\times60=1200$

∴ $x_1+x_2+\cdots+x_{50}=3300$

合併平均為 $\frac{3300}{50}=66$

(3)(4) 低於 68 分至少 15+10=25 人

高於 61 分至少 15+10=25 人

故中位數介於 61～68 分

(5) $8=\sqrt{\frac{x_1^2+\cdots+x_{30}^2}{30}-70^2}$

得 $x_1^2+\cdots+x_{30}^2=30\times4964$

$10=\sqrt{\frac{x_{31}^2+\cdots+x_{50}^2}{20}-60^2}$

得 $x_{31}^2+\cdots+x_{50}^2=20\times3700$

$x_1^2+\cdots+x_{50}^2=$

合併標準差為

$\sqrt{\frac{222920}{50}-66^2}=\sqrt{102.4}>10$

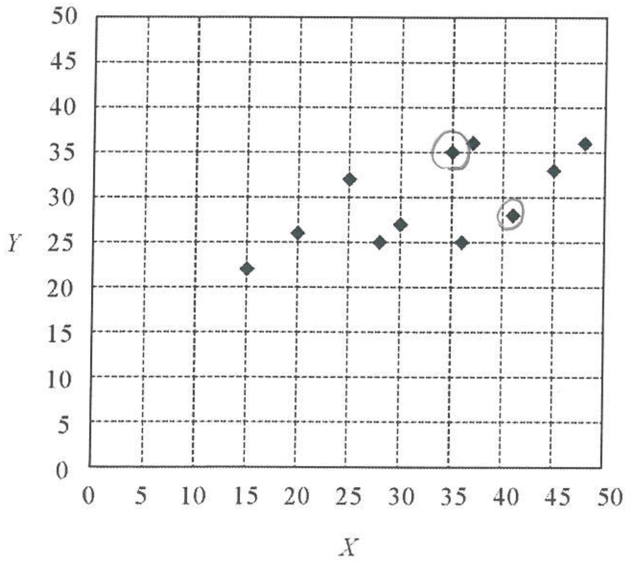
選(2)(4)#

EXAMPLE 10

某次數學測驗分為選擇題與非選擇題兩部分。下列的散佈圖中每個點 (X,Y) 分別代表一位學生於此兩部分的得分，其中 X 表該生選擇題的得分，Y 表該生非選擇題的得分。設 $Z=X+Y$ 為各生在該測驗的總分。共有 11 位學生的得分數據。試問以下哪些選項是正確的？

(1)X 的中位數>Y 的中位數　　(2)X 的標準差>Y 的標準差　　(3)X 的全距>Y 的全距

(4)Z 的中位數=X 的中位數+Y 的中位數



(1) 由左而右第6個數 $x=35$，即 X 中位數為 35
由下而上第6個數 $Y<30$，即 Y 中位數小於 30 (o)

(2) X 約從 15 到 48，Y 約從 22 到 36，X 較分散，標準差較大 (o)

(3) 同(2)，X 的全距較大 (o)

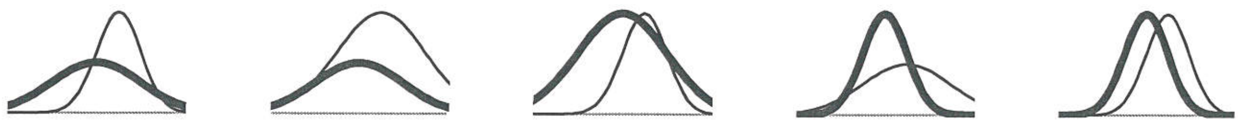
(4) ∵ X, Y 發生在不同點
故 Z 的中位數 ≠ X 的中位數 + Y 的中位數 (×)

選 (1)(2)(3)#

EXAMPLE 11

甲、乙兩校有一樣多的學生參加數學能力測驗，兩校學生測驗成績的分布都很接近常態分布，其中甲校學生的平均分數為 60 分，標準差為 10 分；乙校學生的平均分數為 65 分，標準差為 5 分。若用粗線表示甲校學生成績分布曲線；細線表示乙校學生成績分布曲線，下列哪一個分布圖較正確？

(1)　　(2)　　(3)　　(4)　　(5)



① 學生數相同 → 曲線下面積相同 ② 乙標準差小，較集中 ③ 乙平均大，右邊人多，選 (1)#

EXAMPLE 12

小明參加某次國文、英文、數學、自然、社會五個科目的測驗，每一科的分數均為 0~100 分。已知小明國英數三科的分數分別為 75, 80, 85 分。試問下列哪些選項會讓小明五科成績的平均不低於 80 分且五科標準差不大於 5 分？

(1) 自然 75 分，社會 80 分　　(2) 自然與社會兩科皆 80 分　　(3) 自然與社會的平均 85 分

(4) 自然與社會兩科之和不低於 160 分且兩科差距不超過 10 分

(5) 自然與社會兩科的分數都介於 80 與 82 分之間

(1) $\mu=79$，(×)

(2) $\mu=80$，$\sigma=\sqrt{\frac{5^2+5^2}{5}}=\sqrt{10}<5$ (o)

(3) 若自然 70，社會 100，
則 $\sigma>5$ (×)

(4) 若自然和社會均 100，則 $\sigma>5$ (×)

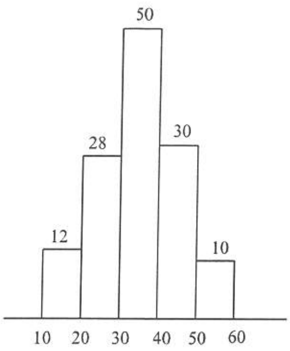
(5) 若自然、社會均 80，則平均 80，故 $\mu>80$
每一科與平均距離，除國文 (5~6)，其餘均小於 5
(皆 0~2 有3科)
故 $\sigma<5$ (o)

選 (2)(5)#

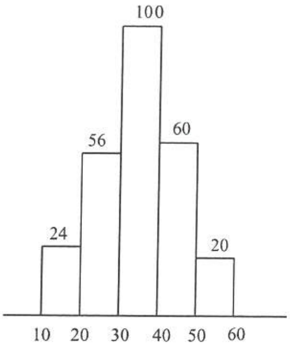
EXAMPLE 13

下列五個直方圖表示的資料，何者之標準差最大？

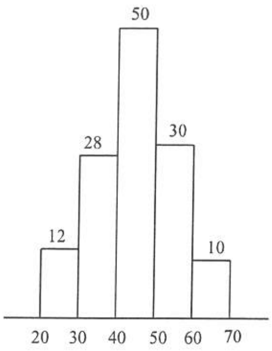
(1)



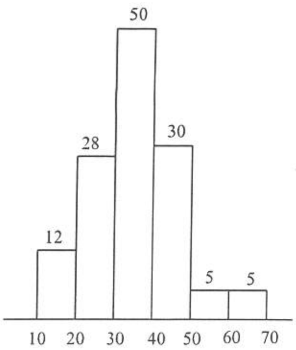
(2)



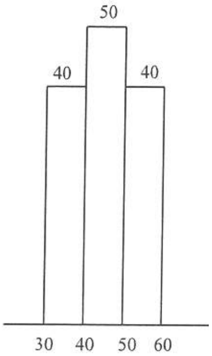
(3)



(4)



(5)

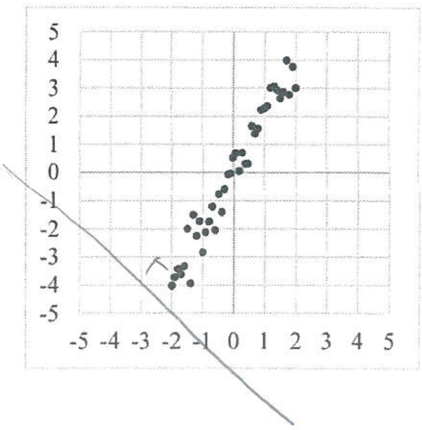


(4) 最分散，故標準差最大.

EXAMPLE 14

在處理二維數據時，有種方法是將數據垂直投影到某一直線，並以該直線為數線，進而了解投影點所成一維數據的變異。下圖的一組二維數據，試問投影到哪一選項的直線，所得之一維投影數據的變異數會是最小？

(1) $y=2x$ (2) $y=-2x$ (3) $y=-x$ (4) $y=\frac{x}{2}$ (5) $y=-x$ $y=\frac{-x}{2}$



點越分散，標準差越大

點越集中，標準差越小

投影在垂直線上，最集中. 資料斜率約為2.

故垂線斜率為 $\frac{-1}{2}$.

選 (5) #

7-2 二維數據分析

1. 相關係數⇒ 具代表性

一筆數據資料有 n 個數據 (x_1,y_1)，(x_2,y_2) ，…，(x_n,y_n) ，則相關係數

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

$$=\frac{(x_1y_1+x_2y_2+\cdots+x_ny_n)-n\mu_x\mu_y}{\sqrt{(x_1^{\,2}+x_2^{\,2}+\cdots+x_n^{\,2})-n\mu_x^{\,2}}\sqrt{(y_1^{\,2}+y_2^{\,2}+\cdots+y_n^{\,2})-n\mu_y^{\,2}}}$$

(1)意義：將 (x_1,y_1)，(x_2,y_2) ，…，(x_n,y_n) 描在散佈圖上的圖形多像 直線 。

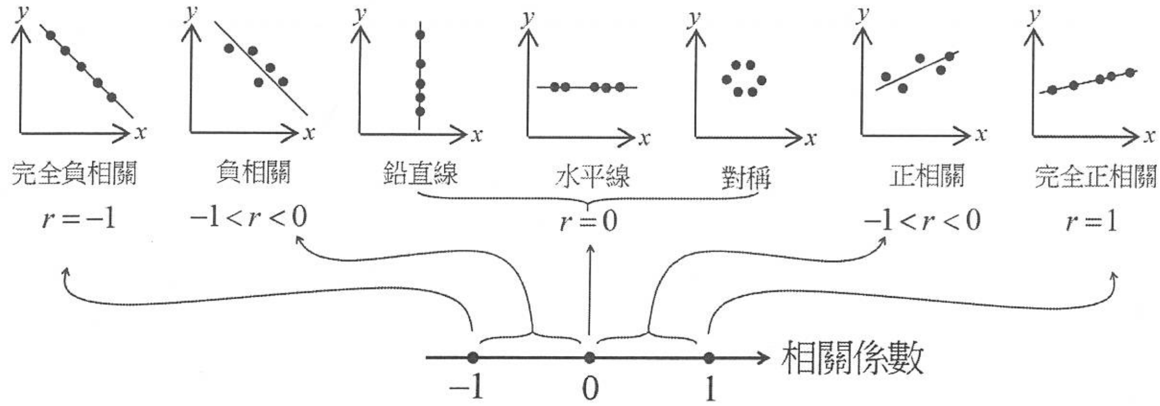
(鉛直線 和 水平線 除外)

(2)性質：r 的範圍： $-1\le r\le 1$

①數值：$|r|$ 越大，表示 X,Y 相關程度 大 ⇒圖形越像直線

②正負：$r>0$ 稱為正相關⇒圖形像斜率為 正 的直線。

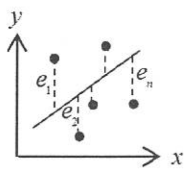
$r<0$ 稱為負相關⇒圖形像斜率為 負 的直線。



2. 迴歸線

兩變數 (X,Y) 的 n 個數據 (x_1,y_1)，(x_2,y_2) ，…，(x_n,y_n) ，其平均分別為 μ_x,μ_y ，標準差分別為 σ_x,σ_y ， X,Y 的相關係數為 r 。散布圖中找出一條最適合的直線 $L: y=ax+b$ 代表 X,Y 的關係，稱此直線為迴歸線。

(1)最小平方法：迴歸線方程式 $y=ax+b$ 滿足殘差和 $e_1^{\,2}+e_2^{\,2}+\cdots+e_n^{\,2}$ 最小。



(2)迴歸線方程式： $y-\mu_y=r\cdot\dfrac{\sigma_y}{\sigma_x}(x-\mu_x)$ （ ①$m=$ $r\cdot\dfrac{\sigma_y}{\sigma_x}$ ②點 (μ_x,μ_y) ）

(3)將數據 (x_i,y_i) 標準化得到新數據 (x_i',y_i')，標準化數據 X',Y' 的迴歸線方程式為 $y'=r\,x'$ 。

3. 相關係數的伸縮與平移

X,Y 的相關係數為 r_{XY}，若 $U=aX+b$ 且 $V=cY+d$，則 U,V 的相關係數為 $r_{UV}=$ $\dfrac{ac}{|ac|}r_{XY}$ 。

EXAMPLE 1

某肥皂廠推出新產品，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)。調查結果如下表，求：

(1) x 和 y 的相關係數　(2) y 對 x 的迴歸線方程式

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

⇒

$x-\mu_x$	-2	-1	0	1	2
$y-\mu_y$	1	2	0	-2	-1

$\mu_x=10,\quad \mu_y=10$

$$r=\frac{-2-2+0-2-2}{\sqrt{4+1+0+1+4}\sqrt{1+4+0+4+1}}=\frac{-8}{10}=\frac{-4}{5}$$

$$\sigma_x=\sqrt{\frac{4+1+0+1+4}{5}}=\sqrt{2},\quad \sigma_y=\sqrt{\frac{1+4+0+4+1}{5}}=\sqrt{2}$$

迴歸線

① 點 $(\mu_x,\mu_y)=(10,10)$

② $m=r\cdot\frac{\sigma_y}{\sigma_x}=\frac{-4}{5}$

$y-10=\frac{-4}{5}(x-10)$

EXAMPLE 2

調查某國家某一年 5 個地區的香煙與肺癌之相關性，所得到的數據為 (x_i,y_i)，$i=1,2,3,4,5$，其中變數 X 表示每人每年香煙消費量(單位：十包)，Y 表示每十萬人死於肺癌的人數。若已計算出下列數值：

$x_1+x_2+x_3+x_4+x_5=135$　　$x_1^2+x_2^2+x_3^2+x_4^2+x_5^2=3661$　　$x_1y_1+x_2y_2+x_3y_3+x_4y_4+x_5y_5=2842$

$y_1+y_2+y_3+y_4+y_5=105$　　$y_1^2+y_2^2+y_3^2+y_4^2+y_5^2=2209$

求：

(1) X 與 Y 的相關係數

(2)若甲每年香菸消費 67(單位：十包)、乙每年香菸消費 35(單位：十包)，則依迴歸直線預測甲罹患肺癌是已罹患肺癌的幾倍？

(1)
$$r=\frac{(x_1-\mu_x)(y_1-\mu_y)+\cdots+(x_5-\mu_x)(y_5-\mu_y)}{\sqrt{(x_1-\mu_x)^2+\cdots+(x_5-\mu_x)^2}\sqrt{(y_1-\mu_y)^2+\cdots+(y_5-\mu_y)^2}}=\frac{(x_1y_1+x_2y_2+\cdots+x_5y_5)-5\cdot\mu_x\mu_y}{\sqrt{(x_1^2+\cdots+x_5^2)-5\mu_x^2}\sqrt{(y_1^2+\cdots+y_5^2)-5\mu_y^2}}$$

$$\mu_x=\frac{x_1+x_2+x_3+x_4+x_5}{5}=27$$

$$\mu_y=\frac{y_1+y_2+y_3+y_4+y_5}{5}=21$$

$$=\frac{2842-5\times27\times21}{\sqrt{3661-5\times27^2}\sqrt{2209-5\times21^2}}=\frac{7}{\sqrt{16}\sqrt{4}}=\frac{7}{8}$$

(2)
$$\sigma_x=\sqrt{\frac{x_1^2+\cdots+x_5^2}{5}-27^2}=\sqrt{\frac{16}{5}}$$

$$\sigma_y=\sqrt{\frac{y_1^2+\cdots+y_5^2}{5}-21^2}=\sqrt{\frac{4}{5}}$$

迴歸線：$y-21=\frac{7}{8}\times\frac{\sqrt{\frac{4}{5}}}{\sqrt{\frac{16}{5}}}(x-27)$，$y-21=\frac{7}{16}(x-27)$

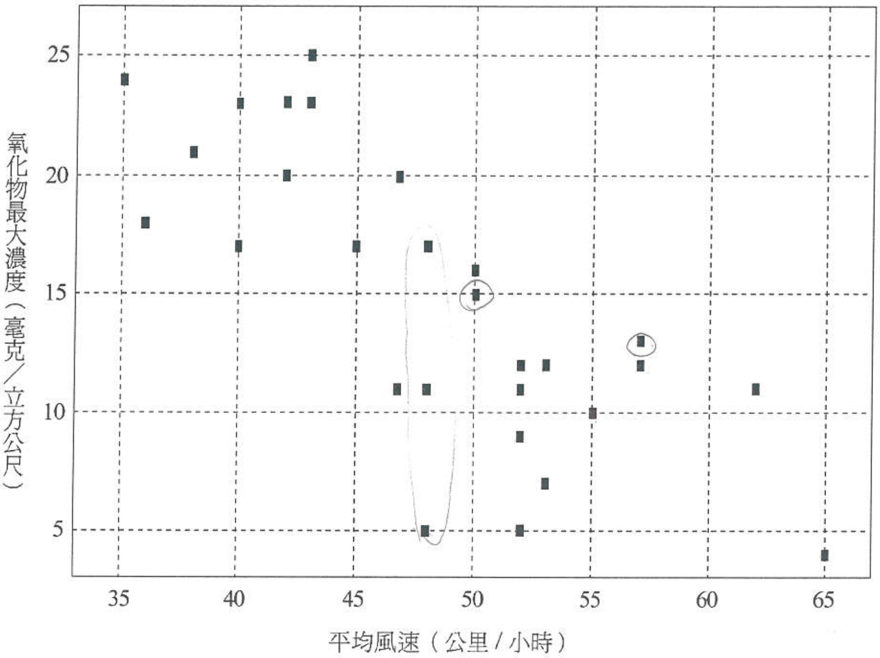
$x_甲=67$ 代入 ⇒ $y_甲-21=\frac{7}{16}\times40$，$y_甲=\frac{77}{2}$

$x_乙=35$ 代入 ⇒ $y_乙-21=\frac{7}{16}\times8$，$y_乙=\frac{49}{2}$

$\frac{y_甲}{y_乙}=\frac{11}{7}$(倍)

EXAMPLE 3

空氣品質會受到污染物排放量及大氣擴散等因素的影響。某一機構為瞭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布圖（見下圖），現根據該圖，可知



(1)此筆資料中，該氧化物最大濃度的標準差大於 15

(2)此筆資料中，該氧化物最大濃度的中位數為 15

(3)此筆資料中，平均風速的中位數介於 45 與 50 間

(4)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

(1) 最大濃度Y值介於 5~25，平均大約 15，故每筆資料與平均距皆小於 15 (×)

(2) 由下往上數第 14, 15 筆資料的平均 (Y=13, 15)，故中位數不是 15 (×)

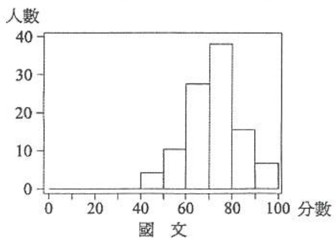
(3) 由左往右數 〃 (X介於 45~50)，故中位數介於 45~50 (○)

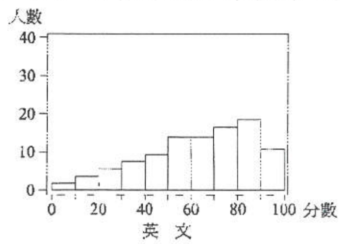
(4) 此資料趨勢直線為 ╲ ，故 $m<0$ (○)

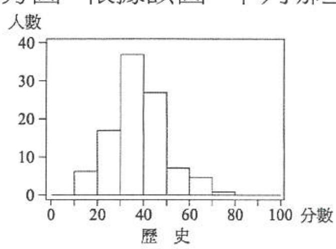
選 (3)(4)

EXAMPLE 4

下圖中為某年級國文、英文、歷史三科成績分布情形的直方圖，根據該圖，下列哪些推論是合理的？







(1) 歷史的平均分數比國文的平均分數低　　(2) 歷史的平均分數最低

(3) 英文的標準差比國文的標準差小　　(4) 英文的標準差最大

(5) 「國文與歷史之相關係數」比「國文與英文之相關係數」高

(1)(2) 平均：歷史在左邊、國文在右邊

故 歷史 < 英文 < 國文

(3)(4) 標準差：英文最分散，故標準差最大

選 (1)(2)(4)

(5) 無法判斷（需散布圖）

EXAMPLE 5

小明參加某次路跑 10 公里組的比賽，下表爲小明手錶所記錄之各公里的完成時間、平均心率及步數：

	完成時間	平均心率	步數
第一公里	5:00	161	990
第二公里	4:50	162	1000
第三公里	4:50	165	1005
第四公里	4:55	162	995
第五公里	4:40	171	1015
第六公里	4:41	170	1005
第七公里	4:35	173	1050
第八公里	4:35	181	1050
第九公里	4:40	171	1050
第十公里	4:34	188	1100

在這 10 公里的比賽過程，請依據上述數據，選出正確的選項。

(1)由每公里的平均心率得知小明最高心率爲 188

(2)小明此次路跑，每步距離的平均小於 1 公尺

(3)每公里完成時間和每公里平均心率的相關係數爲正相關

(4)每公里步數和每公里平均心率的相關係數爲正相關

(5)每公里完成時間和每公里步數的相關係數爲負相關

(1) 此為平均心率 188，最高應大於 188 (X)

(2) 每公里 1000 步即平均每步 1 公尺，由數據知，每公里平均步數大於 1000，故平均每步小於 1 公尺 (O)

(3) 由上而下，時間變少，但平均心率變大，負相關。(X)

(4) 由上而下，步數變多，且平均心率變大，正相關。(O)

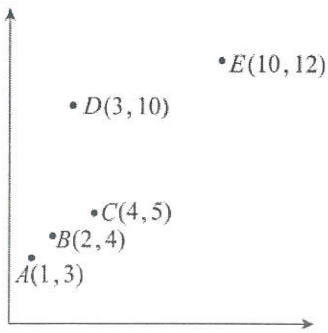
(5) 由上而下，時間變少，但步數變多，負相關。(O)

(2)(4)(5) 為

EXAMPLE 6

如圖所示，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大？

(1)A (2)B (3)C (4)D (5)E



去掉 D 最像直線，選(4)

EXAMPLE 7

根據統計資料，1 月份台北地區的平均氣溫是攝氏 16 度，標準差是攝氏 3.5 度。一般外國朋友比較習慣用華氏溫度來表示冷熱，已知當攝氏溫度為 x 時，華氏溫度為 $y=\frac{9}{5}x+32$；若用華氏溫度表示，則 1 月份台北地區的氣溫平均是華氏______度，標準差是華氏______度。

	℃	℉
平均	16	$\frac{9}{5}\times16+32=60.8$
標準差	3.5	$\frac{9}{5}\times3.5=6.3$

EXAMPLE 8

甲、乙、丙三位同學參加推薦甄選學科能力測驗，五科的成績如下表所示，設 S_1、S_2、S_3 分別代表甲、乙、丙三位同學五科成績的標準差。請仔細觀查表中數據，判斷下列那一選項表示 S_1、S_2、S_3 的大小關係。

	社會	國文	自然	英文	數學
甲	100	70	80	60	50
乙	90	60	70	50	40
丙	80	56	64	48	40

乙 = 甲 − 10 ⇒ $S_2 = S_1$

丙 = 甲 × 0.8 ⇒ $S_3 = S_1 \times 0.8$

$\therefore S_2 = S_1 > S_3$

EXAMPLE 9

令 X 代表每個高中生平均每天研讀數學的時間(以小時計)，則 $W = 7(24 - X)$ 代表每個高中生平均每週花在研讀數學以外的時間。令 Y 代表每個高中生數學學科能力測驗的成績。設 X,Y 之相關係數為 R_{XY}，W,Y 之相關係數為 R_{WY}，則 R_{XY} 與 R_{WY} 兩數之間的關係，下列選項何者為真？

(1) $R_{WY} = 7(24 - R_{XY})$ (2) $R_{WY} = 7R_{XY}$

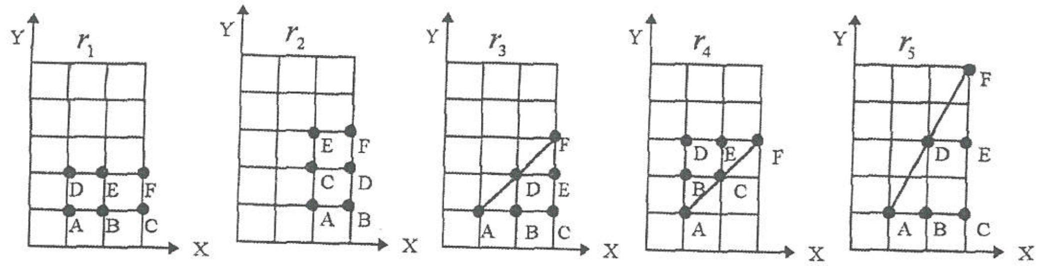
(3) $R_{WY} = -7R_{XY}$ (4) $R_{WY} = R_{XY}$ (5) $R_{WY} = -R_{XY}$

$W = -7x + 168$

$R_{XY} = -R_{WY}$, 選 (5)

EXAMPLE 10

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？

(1) $r_1 = r_2$ (2) $r_2 < r_3$ (3) $r_3 > r_4$ (4) $r_3 < r_5$ (5) $r_4 = r_5$

$r_1 = r_2 = 0$

圖(3) 圖(4) 圖(5)

$(x,y) \longrightarrow (y,x) \longrightarrow (x, 2y-1)$

$R_{xy} = R_{yx} = R_{x,2y-1}$, $r_3 = r_4 = r_5$, 選 (1)(5)

EXAMPLE 11

英國某實驗室研究一金屬圓柱（原高 70.5 英寸）在不同負重下對柱高的影響，其實驗結果如下：

(0, 70.5) (2, 69.4) (4, 68.4) (6, 67.2)

(8, 66.3) (10, 65.5) (12, 64.4)

其中測量單位分別為英噸和英寸。將此筆資料的相關係數記為 r，以最小平方法決定的直線斜率記為 m。現為提供台灣廠商資料，將單位轉換為公噸（1 英噸等於 1.016 公噸）及公分（1 英寸等於 2.54 公分），若單位換算後該資料的相關係數記為 R，以最小平方法決定的直線斜率記為 M。下列關係有哪些是正確的？

(1) $r \cdot m > 0$ (2) $r > 0$ (3) $r = R$ (4) $m = M$ 設

整體而言 x 變大，y 變小 $\therefore r_{xy} < 0$ $\therefore m = r_{xy} \cdot \frac{\sigma_y}{\sigma_x} < 0$

(英噸) (英寸)

X (公噸) Y (公分) ⇒ $\begin{cases} X = 1.016x \\ Y = 2.54y \end{cases}$ ⇒ $R_{XY} = r_{xy} < 0$, $M = R_{XY} \cdot \frac{\sigma_Y}{\sigma_X} = r_{xy} \cdot \frac{2.54\sigma_y}{1.016\sigma_x} < m$ (負多)

選 (1)(3)

EXAMPLE 12

某校高三共有 300 位學生，數學科第一次段考、第二次段考成績分別以 X 、Y 表示，且每位學生的成績用至 100 評分。若這兩次段考數學科成績的相關係數為 0.016，試問下列哪些選項是正確的？

(1) X 與 Y 的相關情形可以用散佈圖表示

(2) 這兩次段考的數學成績適合用直線 $X=a+bY$ 表示 X 與 Y 的相關情形（a,b 為常數，$b\neq 0$）

(3) $X+5$ 與 $Y+5$ 的相關係數仍為 0.016

(4) $10X$ 與 $10Y$ 的相關係數仍為 0.016

(5) 若 $X'=\dfrac{X-\overline{X}}{S_X}$ 、$Y'=\dfrac{Y-\overline{Y}}{S_Y}$ ，其中 $\overline{X}$ 、$\overline{Y}$ 分別為 X 、Y 的平均數，S_X 、S_Y 分別為 X 、Y 的標準差，則 X' 與 Y' 的相關係數仍為 0.016

(1) 散佈圖 適合表示相關情形 (O)

(2) $r=0.016$ 太低，故不適合 (X)

(3) $r_{X+5,\,Y+5}=r_{XY}=0.016$ (O)

(4) $r_{10X,\,10Y}=r_{XY}=0.016$ (O)

(5) $\because S_X, S_Y>0$，故仍為 0.016 (O)

選 (1)(3)(4)(5)