

相 關 分 析

林明滢

台北榮民總醫院醫院感染管制委員會

當筆者仍在學校求學時，對老師所上的統計課只了解到 X^2 -test 及 t-test，提到了相關與迴歸等於是天書一般的『莫宰羊』，一看到那一堆公式，更是我與它們無緣，每逢必跳過，直到近幾年來，才敢與它們有第一類的接觸。其實相關分析與迴歸分析是兩種不一樣的統計方法；迴歸分析是希望以一個自變數來推測另一個應變數的值，具有預測的功能，而相關分析是在探究二個變項間是否有關連存在，不在尋求何者為自變數，何者為應變數，而在探求二個變數間之變化是否有強烈關係，且以相關係數來代表二個變數間關係密切的程度。例如血液中含細菌量高是否血中白血球的數值也高，或反過來說白血球數值高，血中細菌量也高，如果顯著的有此趨勢，我們會說二個變數間有『強相關』，如果例外情形很多，則會認為『相關』很弱；若二者之變化完全無關，則會說二者沒有相關。

於院內感染的資料分析時，我們會去探討很多相關的問題，例如使用消毒劑與降低院內感染有相關，或是患者免疫力與院內感染有相關等諸如此類不同因素間相關性之探討，如兩個變數間是有正相關或者兩者間找不出任何關係存在（零相關）等結果。因此相關分析也會因變項形式之

不同，而有不同的分析方法（請見表一）。本期就介紹一些臨床資料分析時，可能使用的分析方法，由於目前電腦統計軟體相當普遍，筆者就介紹適用資料之範例及其結果之解釋，至於詳細的計算公式，有興趣之讀者請自行參考統計方面之書籍。

表一 兩變項之性質及其適用的相關係數

Y	X	類別變項	次序變項	等距或比率變項
類別變項	ϕ 相關 列聯相關			
次序變項		斯皮爾曼等級相關 肯德爾和諧係數 W 肯德爾 τ 係數		
等距或比率變項	點二系列相關	斯皮爾曼等級相關 肯德爾和諧係數 W 肯德爾 τ 係數	皮爾遜積差相關	

壹、phi 相關 (phi coefficient)

- 一、適用條件：二個變數都是只分二組的類別變項
例如性別（男、女），院內感染情形（是否）
- 二、基本假設：二個變數間無相關存在
- 三、檢定程序：檢定兩組資料間是否有相關 ($\phi=0$) ?

四、檢定統計量：

表二 變數二

	有	無	合計
變數			
有	a	b	a+b
無	c	d	c+d
合計	a+c	b+d	a+b+c+d

$$\phi = \frac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

$$\phi = \sqrt{x^2/N} \text{ 或 } x^2 = N \times \phi^2$$

五、範例：研究者想知道使用導尿管與院內泌尿道感染是否有相關（如表三）

表三 院內感染

	有	無	合計
使用導尿管			
有	13	42	55
無	17	25	42
合計	30	67	97

註：假設資料

$$\phi = \frac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

$$\phi = \frac{42 \times 17 - 13 \times 25}{\sqrt{55 \times 42 \times 30 \times 67}} = 0.181$$

$$\phi = \sqrt{x^2/N}$$

$$x^2 = N \times \phi^2$$

查卡方分佈表 $x^2_{0.95(1)} = 3.841$

$$x^2 = 97(0.181)^2 = 3.17 < 3.841$$

所以 $\phi = 0.181$ 也不顯然小於 0。

未達顯著水準，即是否使用導尿管與院內泌尿道感染無相關

六、決策原則： ϕ 相關係數與 x^2 有密切的關係存在， x^2 值達顯著水準，則 ϕ 值也顯著，若 x^2 值沒有顯著水準，則 ϕ

值也不顯著。

貳、列聯相關

一、適用條件：二個類別變項不只分為兩個類別

二、基本假設：二個變變間無相關存在

三、檢定程序：檢定兩組資料間是否有相關 ($C=0$) ?

四、檢定統計量： $C = \sqrt{x^2/(N+x^2)}$

N 為總個案數

五、範例：調查 634 名工作人員被針頭扎傷者是否較害怕感染 B 型肝炎，其結果如下（表四）

表四 感染 B 型肝炎

	害怕	不知道	不害怕	合計
被扎傷	108	13	122	243
未被扎傷	134	28	229	391
合計	242	41	351	634

註：假設資料

$$C = \sqrt{x^2/(N+x^2)}$$

$$x^2 = 6.717$$

$$p < 0.05$$

$$C = \sqrt{6.717/(634+6.717)} = 0.102$$

x^2 值達顯著 $C \neq 0$ ，即工作人員是否被針頭扎傷與害怕感染 B 型肝炎有相關存在

六、決策原則：扎傷者比較傾向於害怕被感染 B 型肝炎，未被扎傷者較不害怕。用於預測方面，如果我們知道某人被扎傷，其害怕感染 B 型肝炎的可能性為 $108/243 = 44.44\%$ ；

參、斯皮爾曼等級相關 (Spearman rank-order correlation)

一、適用條件：適用於兩個變項都是次序變項的資料，使用在計算兩組等級之間一致程度，如兩個評分者評 N 份試卷，或同一人先後兩次評 N 份試卷。

二、基本假設：二個變數間無相關存在

三、檢定程序：檢定兩組資料間是否有相關？

四、檢定統計量：

$$r_t = \frac{3}{N-1} \left[\frac{4 \sum XY}{N(N+1)} - (N-1) \right]$$

* $\sum XY$ ：等第的乘積和

$$t = \frac{r_t}{\sqrt{(1 - (r_t)^2)/(N - 2)}}$$

五、範例：有一位醫檢師，先後兩次檢查相同十六張尿液抹片觀察尿中白血球含量（個／高倍顯微鏡下），其結果如下（表五）：

表五

抹片編號	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
第一次等第	12	23	63	42	48	32	69	34	71	81	68	35	45	72	92	52
第二次等第	21	36	34	22	56	18	70	20	59	78	96	46	28	47	84	38
	13	9	11	14	6	16	4	15	5	3	1	8	12	7	2	10

請問前後之檢查是否有相關？

註：假設資料

$$r_t = 0.8058$$

$$\text{查表 } t_{14(0.001)} = 2.9768$$

$$t = \frac{r_t}{\sqrt{(1 - (r_t)^2)/(N - 2)}}$$

$$= \frac{0.8058}{\sqrt{(1 - (0.8058)^2)/14}}$$

$$= \frac{0.8058}{0.0253} = 32.196$$

$$t = 32.196 \quad p < 0.001$$

六、決策原則：

得知等級相關係數 (r_t) 為 0.8058， $p < 0.001$ 達顯著結果，表示此位醫檢師的兩次檢查結果甚為接近。

肆、皮爾遜積差相關 (Pearson product-moment correlation)

一、適用條件：兩個變項都是等距或比率變項的資料，是一種最重要和最常用的相關統計法。

二、基本假設：二個連續變項間無線性關係存在？

三、檢定程序：

1. 畫出相關分散圖，觀察資料是否為線性關係？

2. 計算相關係數

3. 檢定母羣體的相關係數是否大於零？

四、檢定統計量：

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \times \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

$$t = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}}$$

五、範例：研究者收集二十位患者的肝功能 ALT 與膽固醇的結果，如下（表六）

表六

患者編號	ALT	膽固醇	患者編號	ALT	膽固醇
1	62	245	11	68	296
2	23	136	12	35	246
3	28	178	13	25	178
4	42	122	14	72	147
5	118	156	15	12	184
6	42	168	16	22	138
7	9	190	17	15	160
8	34	120	18	18	147
9	22	259	19	31	180
10	11	178	20	28	133

註：假設資料

(1)相關分散圖（圖一）

(2)相關係數 r

$$r = \frac{N\sum XY - \sum X\sum Y}{\sqrt{N\sum X^2 - (\sum X)^2} \times \sqrt{N\sum Y^2 - (\sum Y)^2}}$$

$$= 0.1248$$

(3)檢定統計量

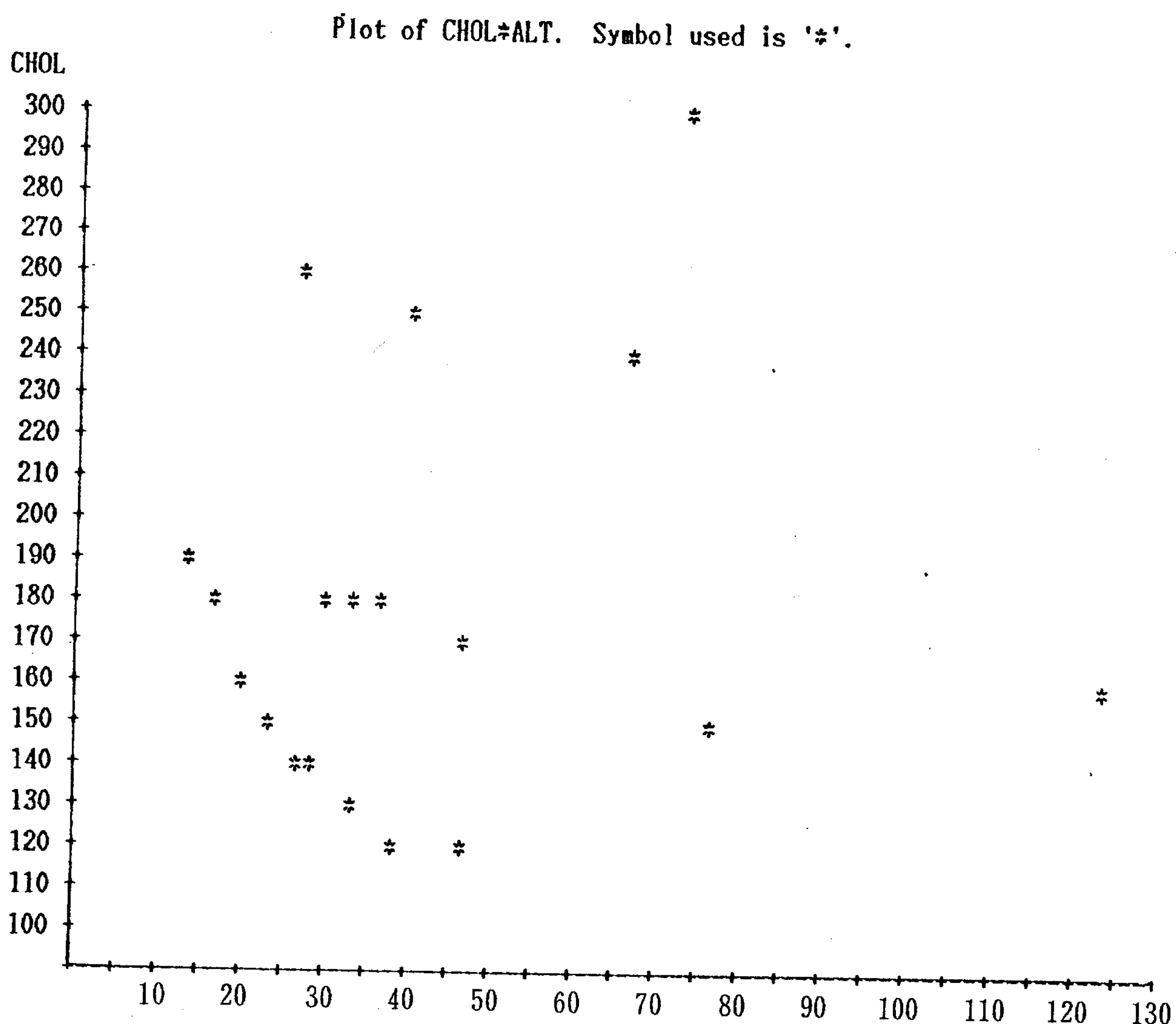
$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.1248 \times \sqrt{20-2}}{\sqrt{1-(0.1248)^2}}$$

$$= \frac{0.5296}{0.9354} = 0.566$$

查表 $t_{18(0.025)} = 2.1009$

統計量 $t = 0.566 < 2.1009$

圖一 相關分散圖



故接受 $H_0: \rho = 0$ ，表示 ALT 值與膽固醇之間，無直線關係存在。

六、決策原則：

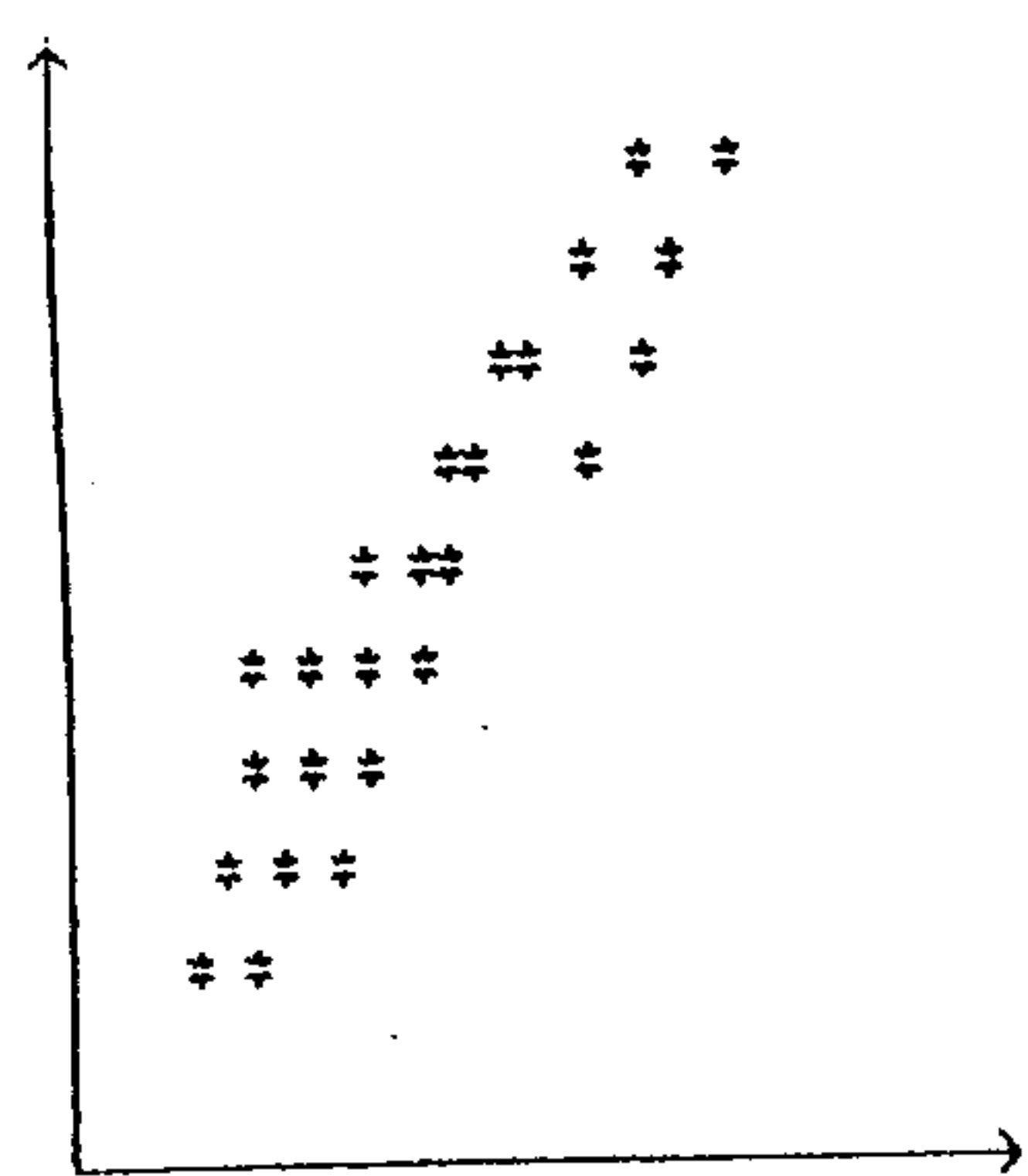
1. 若母羣體的相關係數等於零，表示變項間無線性關係。相關係數未達顯著水準，即表示 ALT 值愈高，膽固醇的值不會愈高。

2. 當相關分散圖畫出後，接著就計算相關係數。相關係數是表示二個變項之線性關係強度之指標，其值介於 +1 到 -1。

(1) $r=1$ 時，表示 X 與 Y 之間的關係為完全正相關（圖二），隨著一個變項 X 的增加，另一變項 Y 也成直線性的增加。且觀察值完全落於直線上。

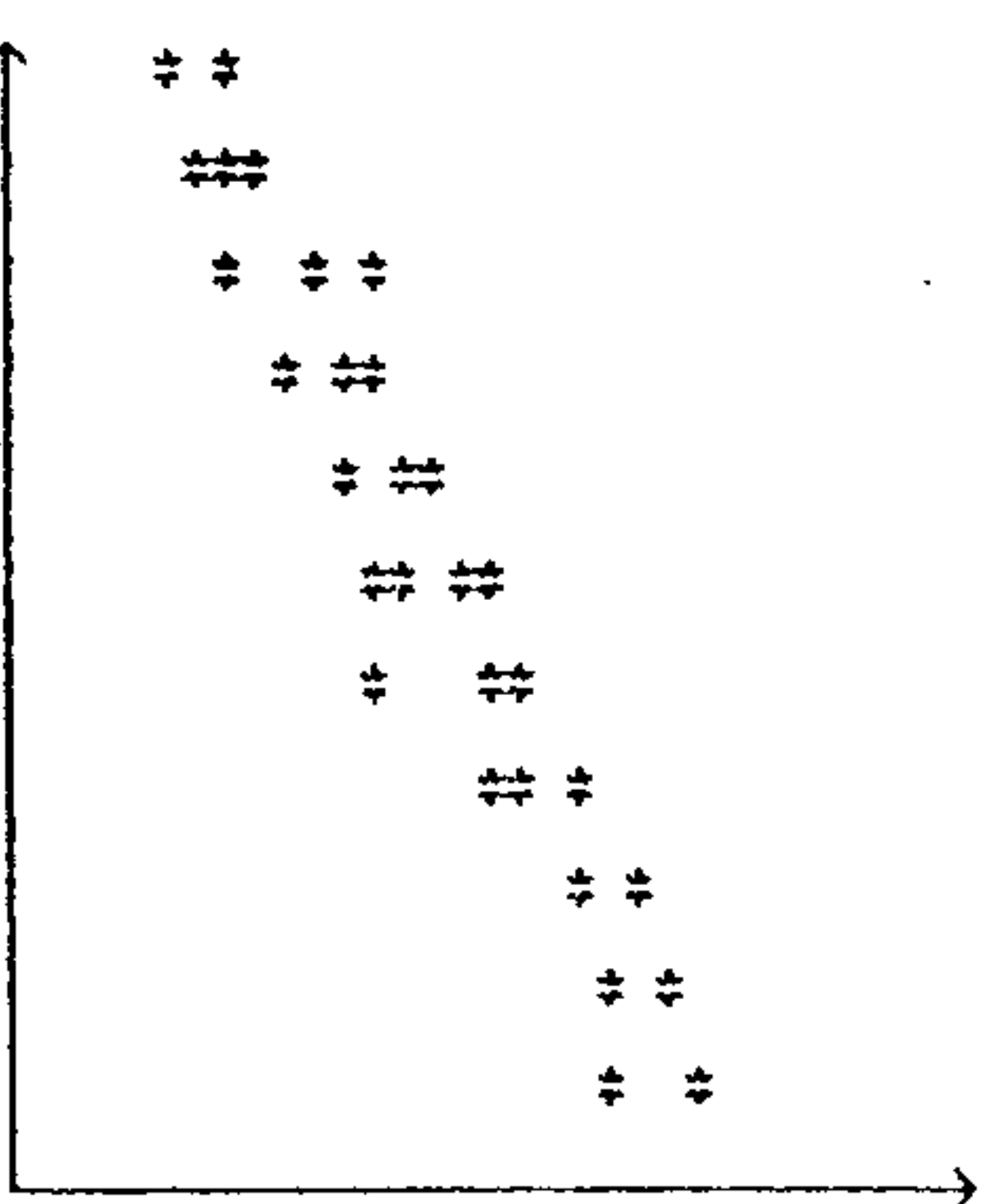
(2) $r=-1$ 時，表示 X 與 Y 之間的關係為完全負相關（圖三），隨著一個變項

圖二



正相關

圖三



負相關

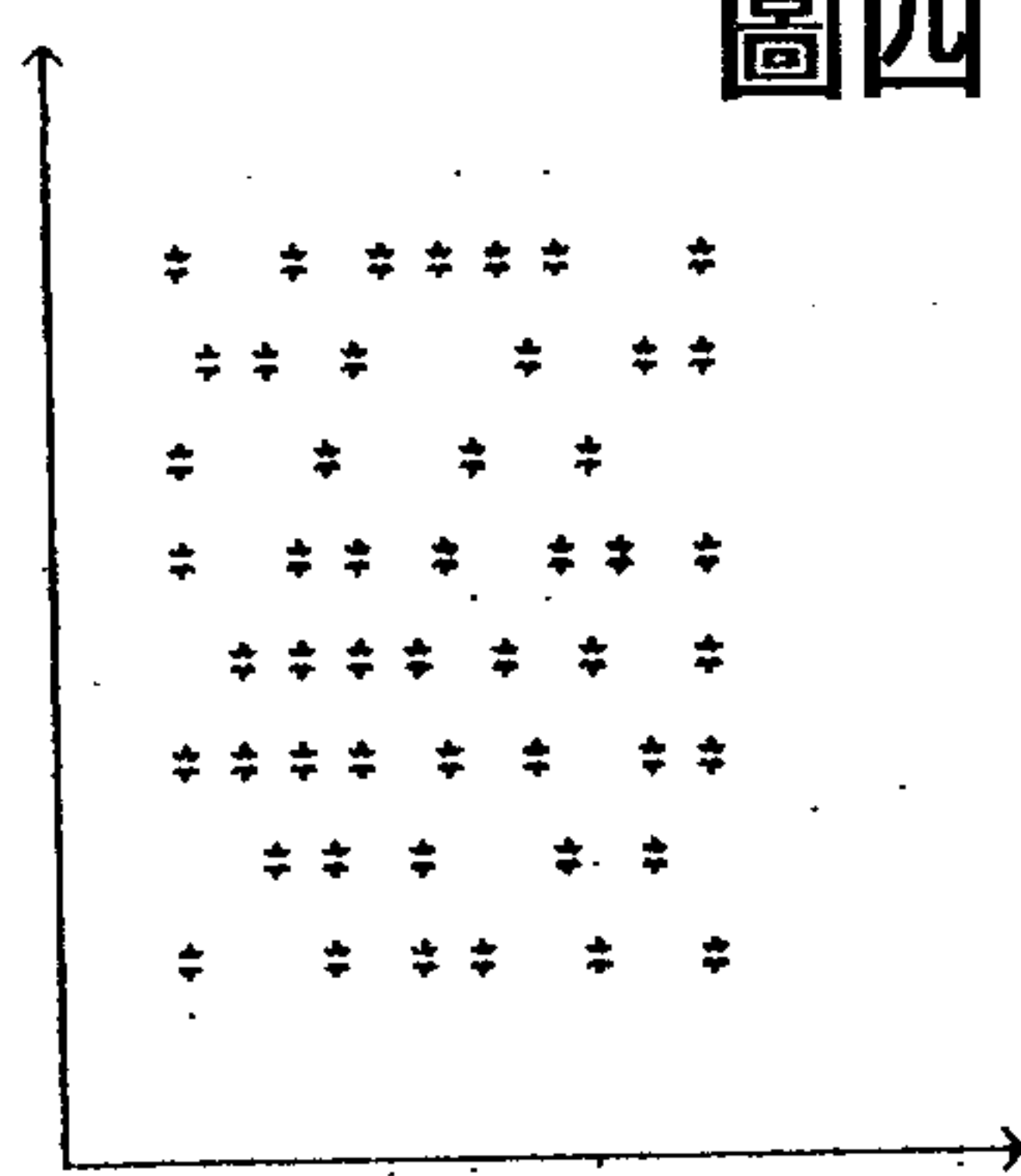
X 的增加，另一變項 Y 也成直線性的減少。觀察值完全落於直線上。

(3) $|r| > 0.7$ 時，表示強相關， $|r| < 0.4$ 時，表示弱相關。

(4) $r=0$ 時，表示 X 與 Y 之間完全不相關（圖四、五）

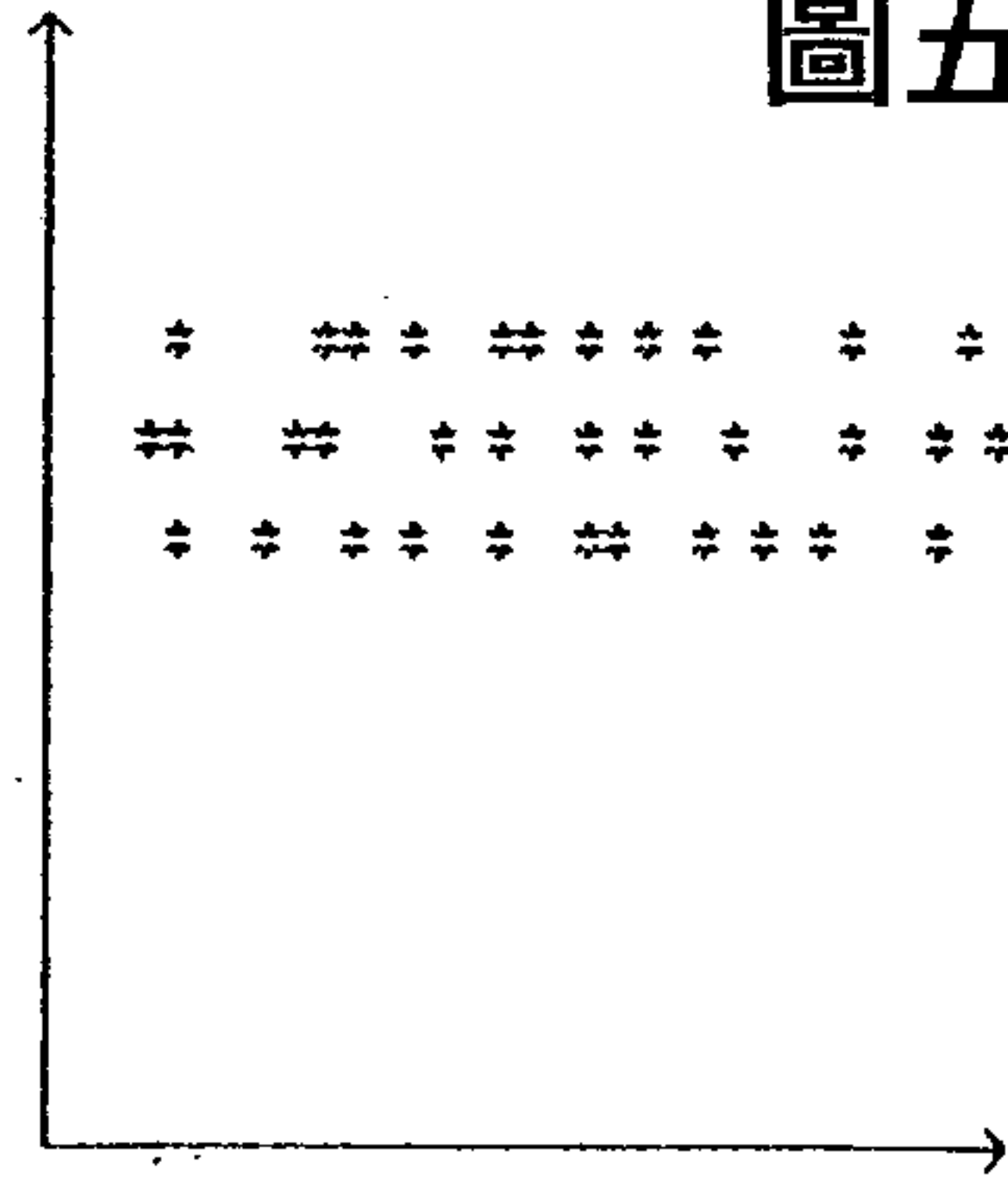
(5) r 很接近零，其相關分散圖可能如圖六、七，此時之資料需轉換。

圖四



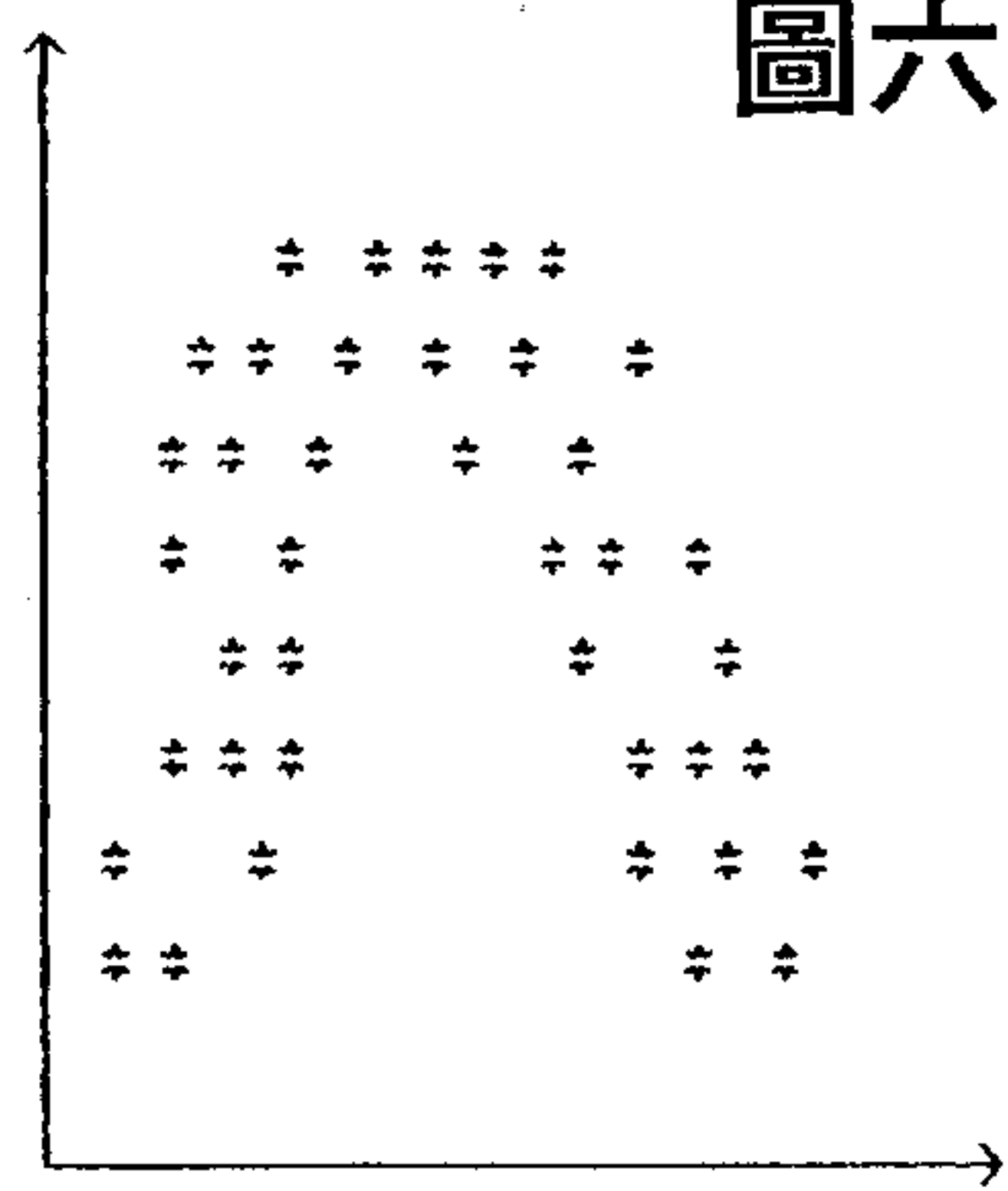
零相關

圖五

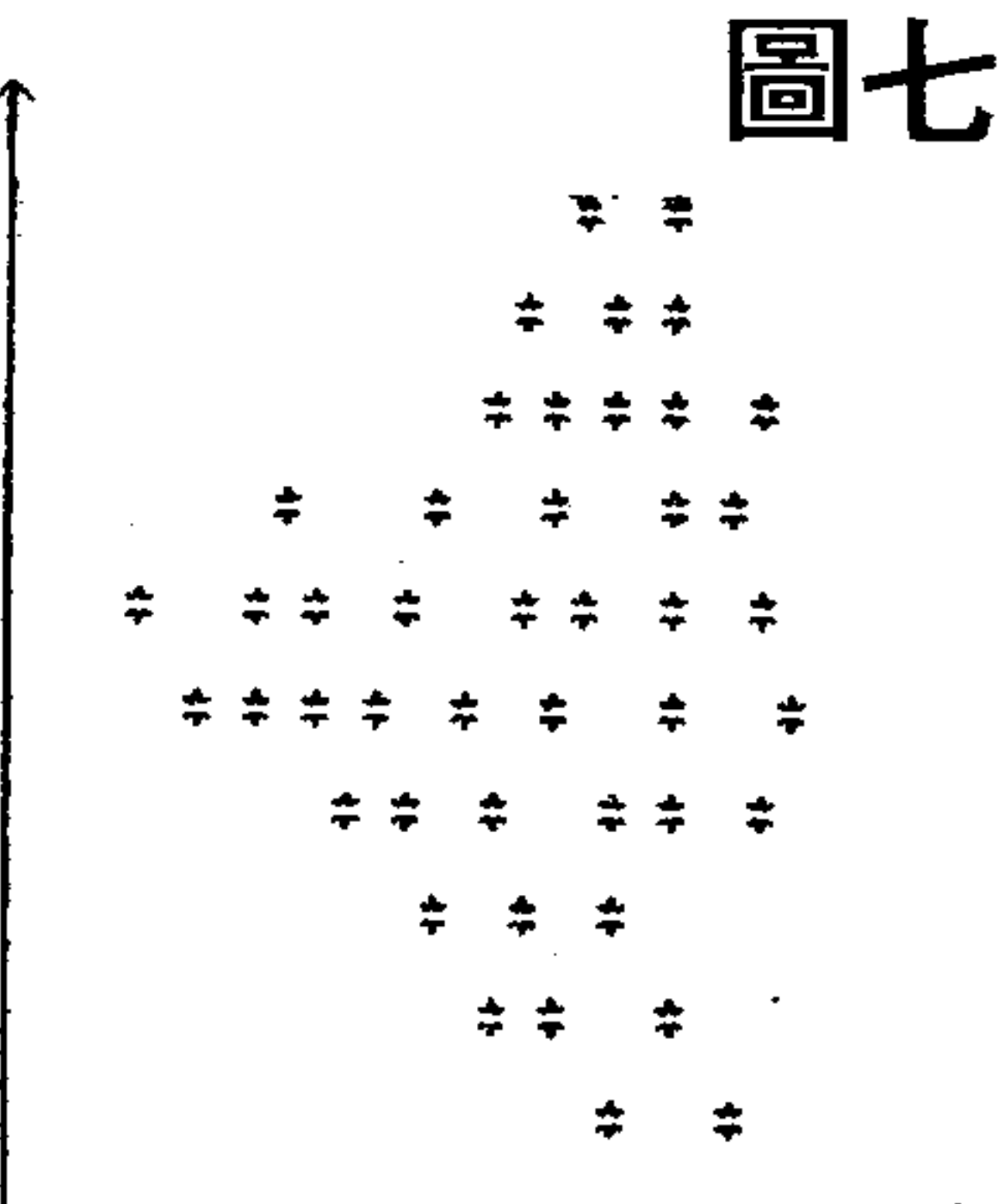


零相關

圖六



圖七



伍、相關檢定的注意事項及限制

1. 若只以相關係數的大小來判定二個變項間的關係，則會有誤差，如果樣本數太少，即使算出很高的相關係數，也不一定可以說兩變項間有相關存在，因高相關可能只是碰巧遇上罷了。

2. 進行相關檢定時，會受到樣本數的影響，況且相關係數本身並非一種比率變數，也不是等距變數，習慣上只是表示關係密切與否的指標而已。

3. 有相關存在並不意謂有因果關係存在。相關愈大就是表示愈可以從其中一個變項較正確的預測另一變項。