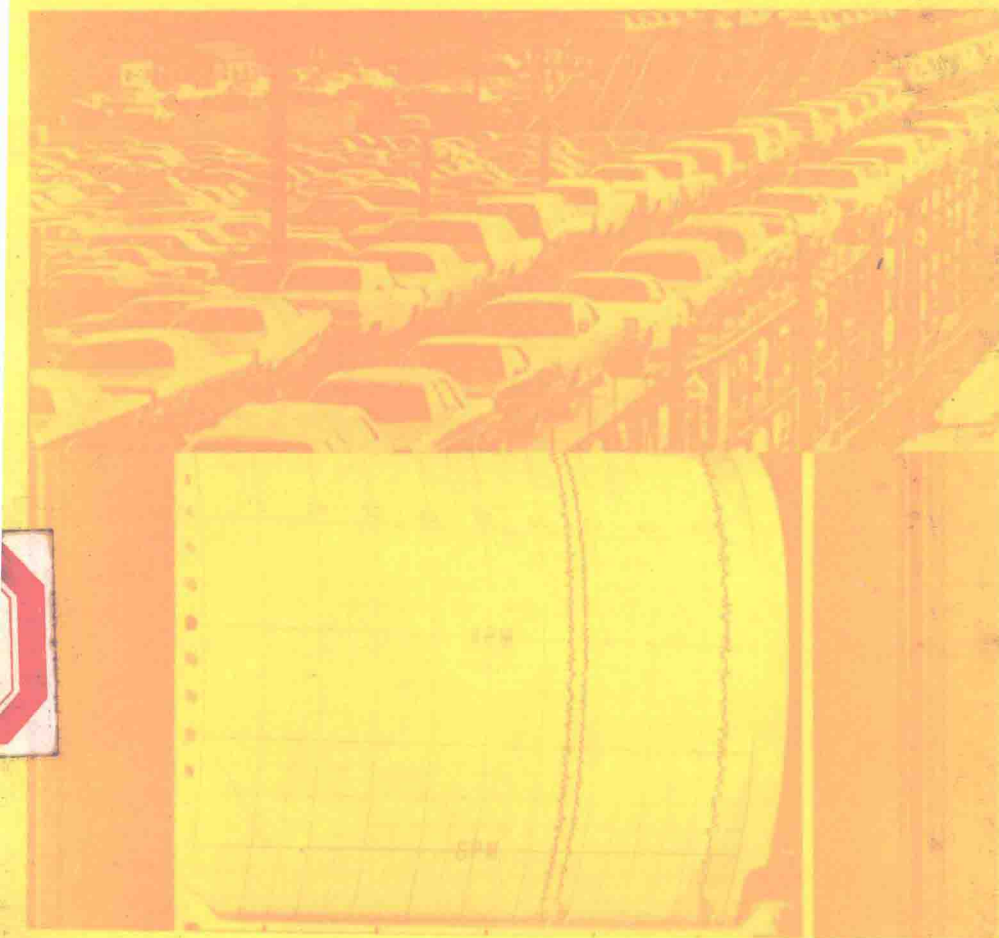


大專教材・品管用書

相關與迴歸分析

附：美國品管工程師資格檢定考試題彙編

蘇芳貴著



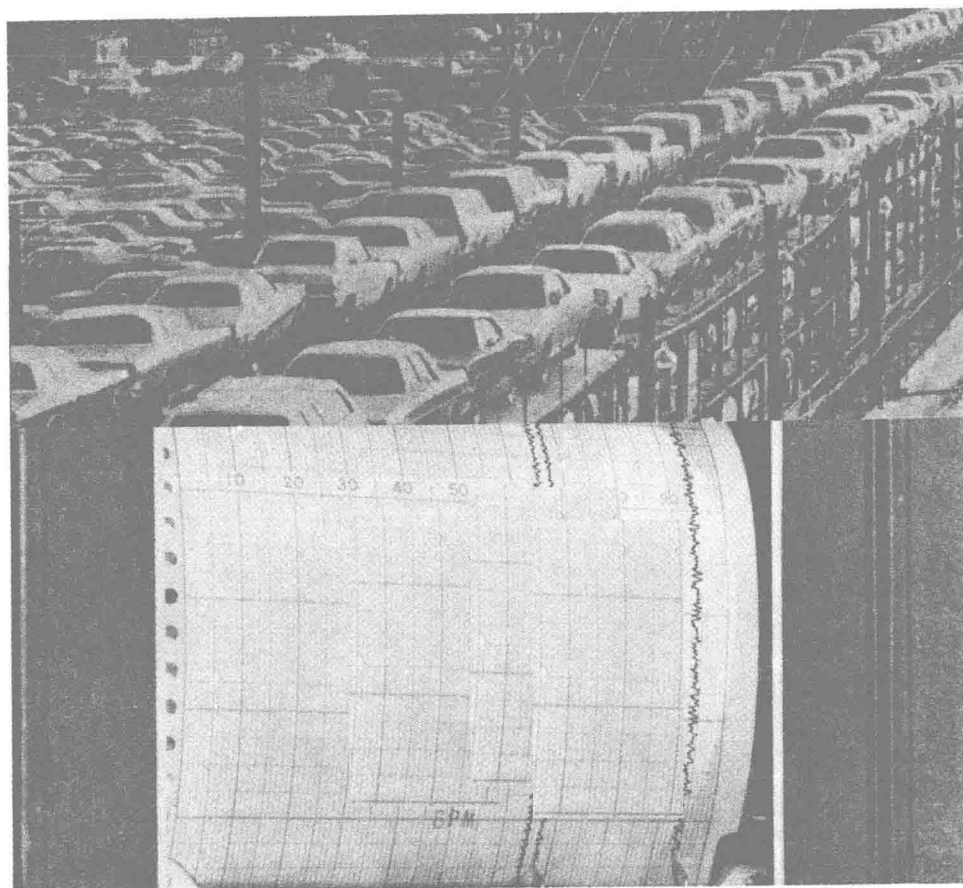
復漢出版社印行

大專教材・品管用書

相關與迴歸分析

附：美國品管工程師資格檢定考試題彙編

蘇芳貴著



復漢出版社印行

中華民國七十二年十月出版

相關與迴歸分析

版權所有
翻印必究

平裝 二〇〇元
精裝 二〇四元

著者：蘇芳貴

出版者：復漢出版社

地址：台南市德光街六五一號

郵政劃撥三一五九一號

發行人：沈岳林

印刷者：國發印刷廠

本社業經行政院新聞局核准登記局版台業字第〇四〇二號

序

管理科學導引人類走向高度開發的科技時代，統計相關與迴歸分析是其中一門學問。目前已廣泛的應用在各方面的經營管理實務上，諸如：製造過程中目標特性的要因分析，影響品質特性的因素探討，實驗計劃配置因素的選定，非破壞型代用特性的研究，電腦製程前饋控制的變數分析，市場行銷預測，營運決策分析等等。上述各項研究工作如果僅靠直覺或經驗的判斷，將承當太多的冒險，而本書所介紹的相關迴歸分析必能使研究結果可靠度和再現性大為提高。

編者以多年來從事公、民營工廠品管實務的經驗以及在中華民國品質管制學會高雄分會負責品管高級班本課程講授的研究心得，相互印證後編纂本書，全書內容儘量兼顧實務和理論，並提供有關的電腦設計程式，以符合社會進步的趨勢和工作者的需要。

此外，本書並提供「美國品管工程師資格檢定考試之問題編譯資料及題解」，將可幫助讀者參加國內品管師和品管工程師的資格檢定考試。（目前由品管學會主辦）該項原文資料係由編者之母校——中原大學盧淵源主任所提供，經由編者和幾位學生合力將其翻譯整理，謹此向上述有關人員致萬分的謝意！

編者才疏學淺，惟願以反哺社會的心情，奉獻自己棉薄之力，希望能給我們的國家多一份管理科學的潛力。公餘之暇，匆促成書，難免有錯誤和不當之處，尚祈賢達先進不吝惠予指正！

編者 謹識

中華民國 72 年 6 月

相關分析與迴歸分析

(Correlation analysis and Regression analysis)

概 述

宇宙間許多事物都在無形中遵循著某些準則反覆的重演，就像人類一直在重蹈覆轍，屢犯錯誤中過日子。高斯（Guass）的試驗似乎也暗示了人類這種無奈！重複的測定工作中，每次的測定記錄都不同，即每次都有誤差；雖然如此，這些錯誤都是有規律可循的。因此，人類利用累積下來的智慧，運用這些規律，避免造成太大的錯誤。而這些規律的探討和模式的建立，其運用的技巧之中，相關分析和迴歸分析是一種簡易而實用的技術。

目 錄

第 1 章 相關分析的觀念	1
1.1 相關分析的意義	1
1.2 相關分析的用途	1
1.3 問題解析步驟	2
第 2 章 直線相關	5
2.1 散佈圖	5
2.2 相關係數	12
2.3 相關表	16
2.4 相關係數的檢定	20
2.5 相關係數的推定	33
2.6 管制圖相關性的判讀	35
2.7 等級相關係數及其分析	36
第 3 章 直線迴歸分析	41
3.1 迴歸分析的意義	41
3.2 直線迴歸方程式或模式	41
3.3 迴歸方程式的估計	42
3.4 由 Y 推定 X 的迴歸式	46
3.5 推定標準誤	49
3.6 迴歸分析之推定與檢定	51
3.7 迴歸分析中的變異數分析	66
第 4 章 模式適合度的探討	69
4.1 剩餘值及對模式適合度的判識	69

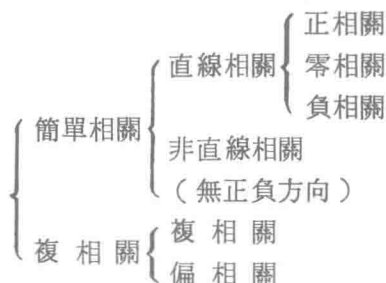
4.2	剩餘值的檢定	71
4.3	線型的 F 檢定	73
4.4	模式的矯正及判斷	77
第5章	高次迴歸、複相關與偏相關	83
5.1	二次拋物線的迴歸方程式	83
5.2	二次拋物線的迴歸分析	85
5.3	複相關	90
5.4	偏相關係數	95
第6章	矩陣的運用	97
6.1	矩陣的定義	97
6.2	矩陣的運算	99
6.3	迴歸分析之矩陣表示法	103
第7章	多元迴歸分析及電腦程式分析	114
7.1	迴歸平面方程式	114
7.2	三元以上迴歸模式的建立和分析	120
7.3	多元迴歸式的電腦程式之應用實例	129
第8章	模式中自變數的選擇	133
8.1	自變數間的相關性	133
8.2	自變數選擇的方法	134
8.3	模式的建立和應用	136
美國品管工程師資格檢定考試問題彙編		137
附 錄		234
附 表		237

第1章 相關分析的觀念

1.1 相關分析的意義

研究兩個變數之間或兩個以上變數之間的關係，稱為相關。譬如：人之身高與體重，其身高較長者，往往體重也較多；其他在工廠製程上，例如反應溫度和產量的關係，觸媒活性和產率的關係，金屬中碳含量和硬度的關係。

相關之種類可區分為：



惟吾人進行探討兩個變數或兩個變數以上之相關性時，所搜集之數據是樣本數據，無法完全代表羣體的現象，故必須利用樣本的相關特性值檢定羣體相關特性是否也和樣本一致，檢定結果有顯著性，再推定出羣體的相關特性。這就是相關分析所研究的範疇。

1.2 相關分析的用途

1.2.1 找尋問題的重點

當有問題發生的時候，通常我們都知道首先探討出各種可能的原因（變數），可是當分析者搜集一切可能原因之後，如何掌握出真正的一、兩項主要原因？這就必須利用到相關分析，以了解到真正與問題相關性較強的有那項原因。針對這些原因先著手追蹤，則可免除許多無效的

分析時間，而達到事半功倍的效果。

1.2.2 製程控制及預防

例如在製程中吾人已確定觸媒活性和選擇性是某一種反應效率的關鍵所在，則觸媒活性和選擇性便應當列入追蹤管制，以確保反應的效率正常；而預先防患這兩項原因造成不利影響。同時由製程分析（Trouble shooting）實驗計劃法的最佳條件研究中，吾人可獲得這些相關性較強的原因（變數），其最佳控制水準在何種範圍，以這些寶貴的情報更可維持製程在最佳的生產狀況。

1.2.3 減少檢驗成本和製造成本

破壞性試驗是不得已的手段，而昂貴的製造原料或零件也是危害企業利潤的重大負擔。如果從許多代用特性和代用品中，吾人確定了某一種代用特性或物料，與我們所欲獲得的結果具有高度的相關性存在，則吾人必然會樂意於利用低成本的代用物料或非破壞性的代用特性以取代目前不經濟的方法。

1.3 問題解析步驟

1.3.1 確定目標特性

1. 問題癥結所在，特性的表達，例如產量、產率、壽命、含量、強度等等。必須能夠以數字表示，而不是含糊的概念。
2. 與專業技術人員交換意見，歸納出一些可能的因素並在理論上事先考慮那些可能存在交互作用（Interaction）。
3. 確定要分析那些和問題有關的因素。並考慮每種因素之單位和情報量的多寡。

1.3.2 搜集數據

這裡必須強調的是成對的或相對的數據，例如在某一化工廠，欲探討製程中某一管制點（AE05）之管制項目 y 含量與前段製程中某一管制點（AE02）之管制項目 x 含量之相關性。則分析人員搜集數據時必

須考慮 AE 02 至 AE 05 之流程時差，如 AE 02 之數據是上午九時取樣分析，則其相對之 y 值（AE 05）必須在流程時差（假設是 1 小時）則 AE 05 之取樣時間最好設定在上午十時取樣分析，以取得 y 值。通常數據最好有 40 對～100 對，太多數據固然是浪費，太少則影響分析的可靠度。

$$\text{通常可靠區間} \propto 1/\sqrt{n} \quad (n: \text{數據數量})$$

例如吾人推定母羣體平均值之可靠區間時，在 95 % 可靠度推定之變異區間為 $\pm 2 \frac{\sigma}{\sqrt{n}}$ ，當 n 太小時，則推定區間必然增大，則推測結果將缺乏意義。

1.3.3 摒除極端值（異常的極大值或極小值）

可利用：(1) $D_4\bar{R}$ 摒除極端值。

(2) 極值檢定法。（參考附錄(一)“Process Quality Control”—Ellis R. Ott, Dixon 檢定）

若造成極端值之原因未知或尚未消除，則不能任意刪除。

1.3.4 繪製散佈圖（Scatter Diagram）

由散佈圖可以瞭解兩種相對數據，在座標軸上點繪而成之散佈點之分佈趨勢，進而判定有無層別（stratification）的必要，同時亦可判定有無上述之極端值（outlier）。

從散佈點之分佈範圍，可以初步瞭解兩種變數間相關性的強弱和方向。

1.3.5 計算相關係數（Coefficient of correlation）

由相關係數可判定兩種變數之間相關程度的大小和方向（正相關或負相關）。

1.3.6 相關分析

由樣本相關係數之分配特性來檢定和推定羣體相關係數。亦即由樣

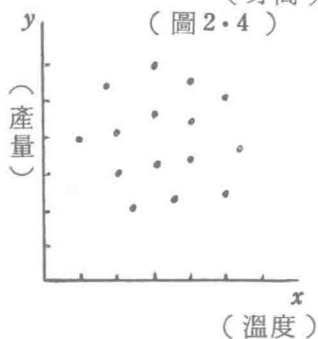
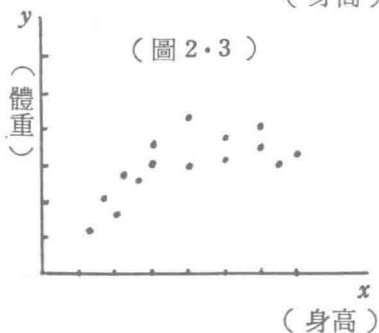
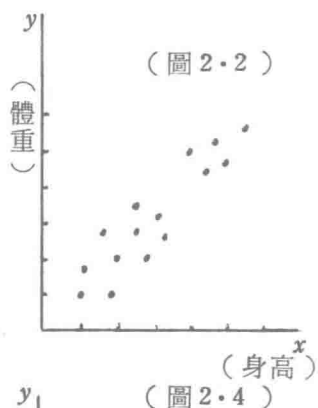
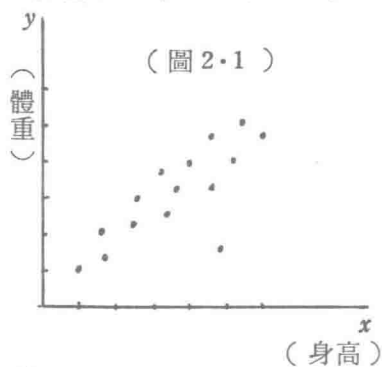
本數據之相關性去瞭解整個製程或羣體的相關性。

第2章 直線相關

2.1 散佈圖 (Scatter diagram)

2.1.1 散佈圖的意義

從所搜集的成對數據 (x_i, y_i)，點繪在座標軸上，則可檢視出有無極端值，(如圖 2.1)，有無離島現象，有無層別的必要(如圖 2.2)，有無非直線的變化(如圖 2.3)，進而檢視出兩變數間概略的相關性大小和方向(x 增加， y 亦增加，或 x 增加， y 減小)。(如圖 2.1，2.3，2.4)



2.1.2 散佈圖的檢討

1. 從圖 2.1，吾人直覺的極易看出有一點在同樣 x_i 的橫座標軸上，顯得格外偏低，其原因值得追查，例如該羣學生體重中，某生在同樣身高學生中顯得體重最輕而又低於身高應有的體重變異範圍，則顯然該生的體重必有問題，否則就是原始數據記錄有誤。（工廠的數據這種現象經常會發生），當該點的異常原因掌握確實之後，便可刪除該點的數據，以免影響統計結果。
2. 從圖 2.2，吾人也極容易發現到所謂的離島型的散佈圖，在這個例子中，追查原因結果發現該羣學生中，有五位是留級的學生，則分析人員可以視情況予以刪除或分開解析、或合併解析。
3. 從圖 2.3，吾人已覺察到這是一種非直線型的散佈圖，亦即在某種身高以上，身高和體重的關係就不成正相關性（即 x 增加， y 亦增加），反而略有下降的趨勢，因此，這 15 對數據不能以直線相關來解析。除非劃分 x 的定義區間（ $\leq x_i$ ， $> x_i$ ）而分別以直線相關來解析。
4. 圖 2.4 顯示， x 和 y 之間的相關性極小，相關分析結果也必然是無相關或極弱的相關性。

2.1.3 散佈圖的作法

一般以橫軸表示原因（或因素、自變數），縱軸表示結果（或因變數）。點繪步驟如下：

1. 搜集成對之數據，40 ~ 100 組，整理成數據表，如下表：

No.	y	x	No.	y	x
1	2	10	7	4	13
2	2	11	8	5	14
3	3	14	9	6	14
4	2	10	10	6	15
5	3	12	⋮	⋮	⋮
6	2	9	⋮	⋮	⋮

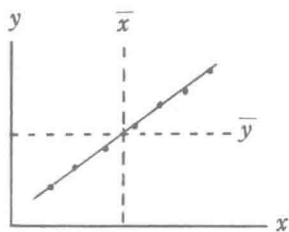
2. 在方格紙上以 x 軸為橫軸， y 軸為縱軸，依據 x 值和 y 值之有效位數決定 x 軸和 y 軸之刻劃間隔，再將各個 (x_i, y_i) 值點繪。
3. 重複出現的點，以 “ $\odot$ ” “ $\odot$ ” 分別表示其該點之重複兩次或三次。

2.1.4 散佈圖的看法

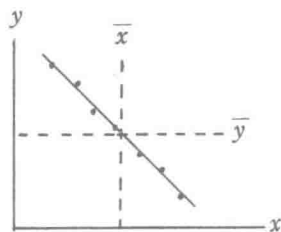
(一) 判別相關的大小和方向

1. 完全正相關和完全負相關

兩組變數數據大小之變化完全成等比例，亦即散佈點完全分佈在一條直線上，而且分佈在第 I 象限和第 III 象限。如圖(a)者，稱 x 和 y 兩變數呈完全正相關。若該直線分佈在第 II 象限和第 IV 象限，如圖(b)者，稱 x 和 y 兩變數呈完全負相關。



圖(a)



圖(b)

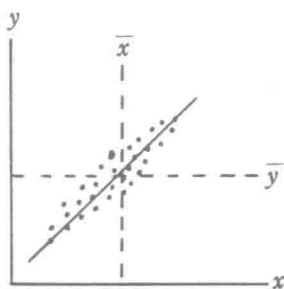
2. 高度正相關和高度負相關

兩組變數數據大小之變化近似成等比例，惟不完全排列成斜直線而略呈較扁的橢圓形的散佈面，而且分佈在第 I、III 象限之點比第 II、IV 象限之點出現更多，如圖(c)者，稱 x 和 y 兩變數呈高度正相關，若分佈在第 II、IV 象限之點出現更多時，如圖(d)者，稱 x 和 y 兩變數呈高度負相關。

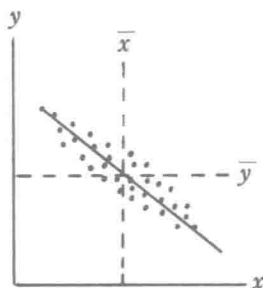
3. 中度正相關和中度負相關

兩組變數數據大小之變化仍近似成等比例，較多散佈點偏離斜直線而略呈橢圓形的散佈面，而且分佈在第 I、III 象限之點比第 II、IV 象限之點出現較多，如圖(e)者，稱 x 和 y 兩變數呈中度正相關。若分佈在第 II、IV 象限之點比第 I、III 象限之點出現較多時，如圖(f)，稱

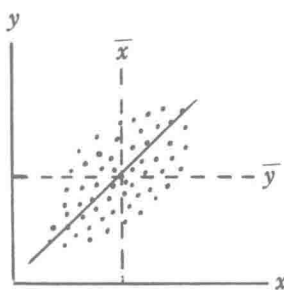
x 和 y 兩變數呈中度負相關。



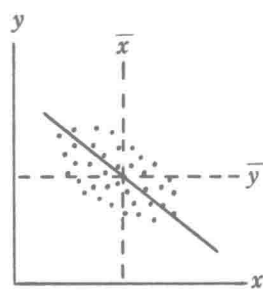
圖(c)



圖(d)



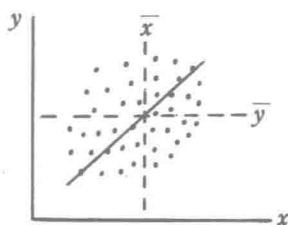
圖(e)



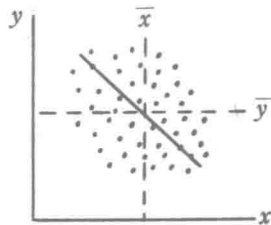
圖(f)

4. 低度正相關和低度負相關

兩組變數數據大小之變化已不成等比例，散佈點偏離斜直線更多，而且第Ⅰ、Ⅲ象限之點比第Ⅱ、Ⅳ象限之點出現略多，如圖(g)者，稱 x 和 y 兩變數呈低度正相關，若分佈在第Ⅱ、Ⅳ象限之點比第Ⅰ、Ⅲ象限之點出現略多，如圖(h)者，稱 x 和 y 兩變數呈低度負相關。



圖(g)

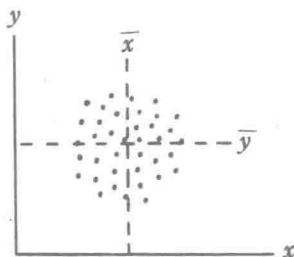


圖(h)

Ⅲ象限之點出現略多，如圖(h)者，稱 x 和 y 兩變數呈低度負相關。

5. 無相關或零相關

兩組變數數據大小完全不成等比例，散佈點均勻分佈在第Ⅰ、Ⅱ、Ⅲ、Ⅳ象限內，如圖(i)。



圖(i)

2.1.5 散佈圖的檢定——符號檢定

從上述散佈圖的看法，對於基層第一線人員，是一種簡便的目視研判方法，惟對於領班以上之人員則應該進一步以檢定的方法來確認目視的研判結果。

這是一種無母數檢定的方法，直接由散佈圖檢定群體（母）相關的大小——群體相關係數（以 ρ 代表）。

(一) 檢定步驟

1. 將散佈圖以 $\bar{x}$ 和 $\bar{y}$ 將散佈點劃分成四個象限，（此時 y 線將散佈點上下均勻分開， x 線將散佈點左右均勻分開）。分別計算第Ⅰ、Ⅲ象限之點數（ $n_1 + n_3$ ），第Ⅱ、Ⅳ象限之點數（ $n_2 + n_4$ ）。
2. 建立無效假設 $H_0 : \rho = 0$ （無相關）
對立假設 $H_1 : \rho \neq 0$ （有相關）
3. 計算 $N = n_1 + n_3 + n_2 + n_4$
4. 決定檢定冒險率 $\alpha = 5\%$ ， $\alpha = 1\%$
5. 查符號檢定表（附表一）

得 $\alpha = 5\%$ 時， N 值的下部界限 λ_L 和上部界限 λ_U 。以及 $\alpha = 1\%$ 時， N 值之 λ'_L 和 λ'_U 。

6. 判定結果



- (1) 若 $n_1 + n_3, n_2 + n_4$ 介於 $\lambda_L \sim \lambda_U$ 之間，則判定 x, y 兩變數間無相關性存在。(即 H_0 成立)
- (2) 若 $n_1 + n_3, n_2 + n_4$ 分別介於 $\lambda'_L \sim \lambda_L$ 和 $\lambda_U \sim \lambda'_U$ 之間，則判定 x, y 兩變數有顯著的相關性。(即 H_1 成立)
- (3) 若 $n_1 + n_3, n_2 + n_4$ 落在 $\lambda'_L \sim \lambda'_U$ 之外，則判定 x, y 兩變數間有極顯著的相關性。

【例】某鋼鐵工廠爲了探討硬度 (y) 和碳含量 (x) 之相關程度，搜集 30 個成對數據，並繪成散佈圖如下：

【解】：

$$(1) H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$(2) N = 30$$

$$n_1 + n_3 = 22$$

$$n_2 + n_4 = 8$$

$$\therefore n_1 + n_3 > n_2 + n_4$$

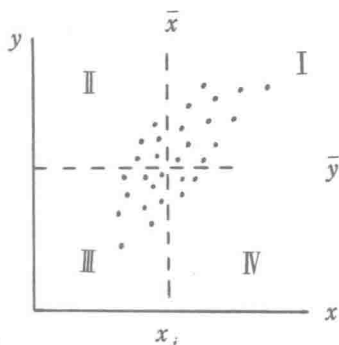
故方向爲正

$$(3) \text{ 取 } \alpha = 5\%, \quad \alpha = 1\%$$

查符號檢定表得知：

$$\lambda_L = 9 \quad \lambda_U = 21 \quad (\alpha = 5\%)$$

$$\lambda'_L = 7 \quad \lambda'_U = 23 \quad (\alpha = 1\%)$$



$$\therefore \lambda'_L < n_2 + n_4 < \lambda_L$$

$$\lambda_U < n_1 + n_3 < \lambda'_U \quad \text{故 } H_1 \text{ 成立}$$

(4) 結論

$\therefore H_1$ 成立 故可判定硬度和碳含量存在顯著的正相關