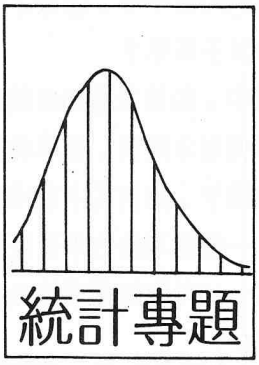


相關的概念

黃登源

第一節 緒 言

在單一變數的資料中使用集中趨勢及變異程度之測定，但在討論兩變數間的關係時，相關的各種指標就非常重要了。考慮兩變數間的關係，要注意到它們是否以同一單位或尺度所測定；或者一變數的變異程度比另一變數大或小；或者它們的關係並不僅是線性關係，而可能是曲線關係。關於這些問題我們將探討一些相關的指標。例如，一變數由另一變數解釋變異的百分比之測定是一種相關的指標。不過，一變數X之變異由另一變數Y之變異所解釋多少與Y變異由X變異之解釋多少可能不一樣。或者我們可以在散佈圖上的點使用一條直線通過這些點來描述做爲相關的指標，不過，在這種情況下，如果一變數用公斤來測定和用公克來測定與相同的另一個變數所描述的趨勢將不會一樣。

當我們提到相關的時候，就會想到一個變數與另一變數手牽手的連接在一起，換言之，即個體在一變數的次數與在另一變數的次序有相似之處，或個體在兩變數上有相同的地位。像這樣的情形，我們就說在特定的一群對象上之兩變數有好的關係。

本文主要在介紹兩變數間之線性關係有關概念，內容根據Ghiselli等（1981）之著作改寫而成。至於相關理論則屬於多變量統計分析範圍，其應用非常廣泛，了解相關概念是進入此一範圍的先決條件。由於此種概念牽涉太廣，不易在此短文敘述完全，僅供讀者有個概略印象而已。

第二節 個體差異之描述

對於種類而言，個體之變異通常是質的特性，但對於次數、數量或程度而言，是量的特性。當我們研究個體差異時，我們必須定義或特殊化我們所關心的性質。由此種定義我們可以發展一序列作業或度量方法使我們可以根據此種性質的個體加以描述。質的描述稱爲分類，而量的描述稱爲度量。度量牽涉到數值的使用，也就是此種數值提供個體量的描述及進一步提供個體的資訊。

2.1 變數之描述

個體之性質可能是光滑、大小或態度等之特性。一些性質可能是描述某一個特定母群體

所有成員或個體，因此所有生物有生命的性質，所有的書有頁數的性質，及所有的魚有住在水裏的性質。另一方面，某一母群體之成員間對於某一特定性質具有差異。因此，某種性質在各個個體之間有差異者，此種性質稱爲變數。形狀是此種性質如石頭具有各種不同之形狀，重量也是此種性質如人類具有不同之重量等。對於某種性質而言有各種方式之個體差異。它們可能在種類上質的差異或在數量上的差異。大學生不僅在主修科目種類之不同也在學習成就上不同。更進一步來說，有各種不同型態的量的變異。在一些情形下，量的變數是不連續的及顯示間斷的，有層次的，如表示「火車車廂數目」之變數就是不連續的，在此分數的車廂是無意義的觀念。另一方面，連續的量的變數如長度，在此個體差異可能非常小也可能非常大。

2.2 變數結構

我們已經定義特徵或性質在個體間有差異者，稱爲變數。在某些情形下此種特徵或性質相當具體。因此若我們有興趣於每個家庭的孩子數，我們可以看、接觸或觀察此種家庭特徵。但在其他情形，特別是心理變數就不具體了。雖然個體很具體，但如智力就不是了。

長度是什麼呢？決不是用來測定的尺，也不是用鉛筆畫出的線，而是兩點間想像中的直線。它是一種觀念。許多人類的非常重要特徵是不具體的，而是需要間接測定及逐步接近我們心目中的變數。事實上似乎是越重要的變數越難於測定。例如，學習態度，工作動機，應變能力及創造力等。

比較上，物理及化學之測定工作較簡單。物理科學對於有興趣的變數利用標準來測定，例如國家度量衡標準，這些觀念因此容易測定並且完全一致性。但心理測定就不是這樣的，對於有興趣的變數之測定或定義是否被接受將會有疑問。

2.3 變數與常數之性質

個體在種類或數量有差異之特徵或性質稱爲變數。但若這些個體之特徵或性質只有一個種類或數量，此種特徵或性質稱爲常數。對全人類而言，性別是變數。但只對男性而言，性別是常數，及只對女性而言，性別是常數。在數學測驗上不同學生得到不同分數，反應在分數上之數學能力是變數。但是，所有得 42 分之學生的數學能力是常數。

2.4 質與量的變數

所有變數可以分成兩種一般型態：質的變數和量的變數。當變數是質的特性時，是指個體在種類上之差異；及當變數是量的特性時，是指個體在次數、程度或數量上之差異。對於質的變數而言，個體之類別是沒有次序的，但數量變數之分類是有次序的。

職業是質的變數之例子。我們可以將工作者分爲經理、銷售員、記帳員、服務員或操作員等。在此種分類系統中沒有自然次序。對於量的變數，有分類的自然次序因爲不同類別表示不同程度、數量或次數。

2.5 量的變數之型態

量的變數可以分爲兩種型態：順序變數和尺度變數。順序變數僅提供個體之次序，但是尺度變數提供特徵的次數、程度或數量。在家庭中孩子的位置如第一次出生、第二次出生等，這是順序變數的例子。人的重量如張三體重 60 公斤及每年收入 50 萬元等這些是尺度變數的例子。後面這些指出順序和數量也指出個體間的差異。

2.6 順序變數的型態

順序變數提供一序列分離或離散的類別，不過與質的變數不同之處在於順序變數是有次序的。這一序列類別不需表示它們的特徵或性質的差距，而僅是多或少，並沒有特別指出差距之數量。

順序變數不僅可以描述個體之順序，也可以描述群體之順序。例如教師可以把學生分成有興趣、無所謂及沒有興趣三群學生，以「對數學之興趣」爲變數來排序。

2.7 尺度變數的型態

尺度變數可以分爲差距尺度和等比尺度。等比尺度之絕對零已知，而差距尺度的絕對零未知。

例如，我們可以說一個人身高150公分是另一個身高 75 公分的人的兩倍，因爲量身高的直尺有絕對零點。但是假設我們測定算術能力，我們不能說一個解不出一個問題的人的算術能力爲零，因爲此人可能解出更簡單的問題。因此在一個測驗中得零分，並不表示絕對零或完全沒有算術能力。

2.8 依據變數決定個體差異

當我們研究個體差異時，首先必須定義我們有興趣的變數。由許多我們所關心的個體所顯示的特徵和性質，我們取出一個特別性質描述與區別個體間的差異。我們的定義應該指出變數之型態是質或量，如果是量，是什麼型態。變數定義好之後，我們設計一序列作業或測定方法觀察個體間的相似性和差異。這些作業牽涉到使用一序列規則規定使用程序和儀器，而且提供個體之類別使我們可以依據這些變數描述。

2.9 定義變數

我們定義變數是爲幫助了解性質之本質及作爲發展作業之基礎以描述該性質之個體差異。定義變數不是一件容易的工作，我們以下討論一些定義變數之注意事項。

(1)變數的定義越清楚和越明確越有用。

(2)選擇變數適當的名字給合理的代表性。

(3)變數是理論與智識的函數。

2.10 在分類、排序與測定上使用數據

對測定而言，數據有兩種重要特徵。第一，它們提供方法依照某種特徵之程度有系統方法將個體分類或安排。例如，在數學測驗上得 42 分學生在同一類別，得 43 分者在另一類。

數據之第二種特徵是它們可以由算術程序作業和組合使它們能更精確描述或表達其他意義。例如，對於同一個體使用同一工具重複測定，求出這些重複測定值的平均數，在這種更多測定下獲得更精確量的描述，也就是獲得個體能力的更可靠的估計。另外，在四個教師對於小孩的侵略性等級測定，若要得更多測定值，希望由其他教師來測定，這樣我們可以獲得對該個體更好的描述。

(1) 質的變數

當我們處理質的變數時，我們指定個體到兩個或兩個以上類別或組別中之一。通常我們指出類別以表示某種性質，例如，性別有兩種指示法，即男性及女性及變數「在大學主修」，其指示法爲「經濟學」、「歷史」、「心理」與「化學」等。

不過，通常我們使用文字或數字作爲指示。當我們使用數字時，僅用來表示類別，這些數字稱爲類別尺度。

(2) 順序變數

當我們處理順序變數時，根據表示某種性質的次數、數量和程度將個體或群體排序。按

照順序以 1，2，……表示，此種數字稱爲順序尺度。此種尺度只知所佔位置不同而不知它們的差距。

(3) 差距變數

差距變數提供一序列有順序的類別，在連接類別間之個體對某種性質在頻率數量或程度上之差距是可以知道的，如在尺度的數字之間的區間相等或已知，這就是差距尺度。此種尺度可以測量兩點間的距離，例如，華氏溫度計是差距尺度，因爲我們可以將每個溫度加或減並不改變區間的意義。

(4) 等比變數

對於某些變數而言其作業結果會產生等比尺度，此種變數有絕對零，例如年齡變數，15歲是一個很精確的數字，我們可以說30歲的人是15歲的人年齡的兩倍。

2.11 連續尺度及作業

對於連續量的尺度在兩個體間的差距可大可小，理論上此種差距可以變得很小很小一直到差距消失而成同一點。在實際作業上不可能提供連續尺度，事實上僅靠近到符合我們的需要就可以，我們可以給一區間使所有在同一區間的個體給予同一數值。因此當尺度是連續時，指定於個體量的描述不是精確的而是近似值。測定工具可能是粗糙的，也可能是精密的。

第三節 相關之測定

在前面我們描述一變數的測定尺度及其概念。在這一節中我們將針對成對變數（X,Y）探討它們之間的關係。主要考慮此兩變數之測定單位或它們各自的變異程度或者它們的關係不是線性，而是曲線。我們主要討論一種線性關係測度——皮爾遜相關係數（Pearson correlation coefficient）。

上面所提到皮爾遜相關係數不受各自的平均數及變異數的影響，也就是不因測定值的起點及所使用的單位而改變。所得係數僅受各變數本身次序關係及機率分配形狀之影響。也可以說相關係數爲尺度不拘（scale free）。

此種相關係數僅當（X,Y）顯示線性相關時測定相關程度。我們不擔心由樣本資料估計母群體之特徵值問題，因爲我們主要在探討如何提出較合適描述而不在統計推論，所以我們選擇變異數

$$\sigma_x^2 = \frac{1}{n}\Sigma\,(x-\bar{x})^2$$

而不以$n-1$代n之原因在此。雖然以$n-1$代n，其估計母群體變異數將稍有略小偏差（bias），但從較大母群體中取出樣本，其平均值之影響極小。此處之母群體係指針對所考慮主題的對象全體，樣本則指從母群體中取出一部份有代表性的對象觀察所得的結果，由此我們將σ_x^2代表母群體之變異數。機率分配係指所有變數值分佈狀況。

3.1 變數間關係之型態

檢查兩變數間相關的性質，我們可以畫散佈圖，散佈圖可以用來表示二變量聯合次數分配，此圖形是把兩變數值的分佈情形表示出來。

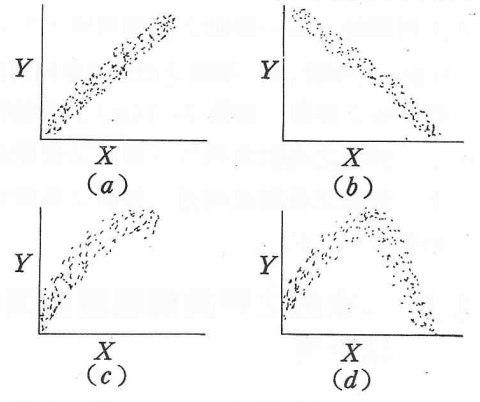


圖 3-1 線性及非線性關係散佈圖
(a)正線性關係 (b)負線性關係
(c)非線性關係 (d)非線性關係

由上圖可以看出兩變數之間有許多不同型態的關係，可能是線性，也可能是非線性；可能高，可能低或中等程度之相關；可能正或負相關。

3.2 線性和非線性相關

因爲線性關係在許多現象間的關係是很好的表示法，通常使用它來測量兩變數間的關係，也有許多關係很接近線性，就以線性表示，但是還有些情況必須用非線性來表示。我們必須在事先觀察決定使用線性或非線性，其方法可以使用畫圖的方法評估使用線性或非線性較合適。

3.3 相關程度

在一些特徵之間有些彼此結合程度很高，有些中等，而有些可能完全無關。當兩種特徵之關係很高，其意義就是它們幾乎由同一因素決定。當相關低時，其意爲兩特徵之個體差異由相當不同因素決定。因此，相關程度是瞭解變數的一種方法。

3.4 正與負相關

兩變數之相關方向可能正或負。當兩變數呈正相關時，則一變數之數值高傾向於結合另一高數值之變數，而一變數之低數值傾向於結合另一低數值之變數。如圖 3-1(a)。負相關則相反，一變數之高數值與另一變數之低數值結合，而一變數之低數值與另一變數之高數值結合，如圖 3-1(b)。

3.5 以變數之變異解釋變數間的結合性

圖 3-2 是兩個變數X及Y的聯合次數分配，在散佈圖左及下分別表示在Y及X上之邊際分配。

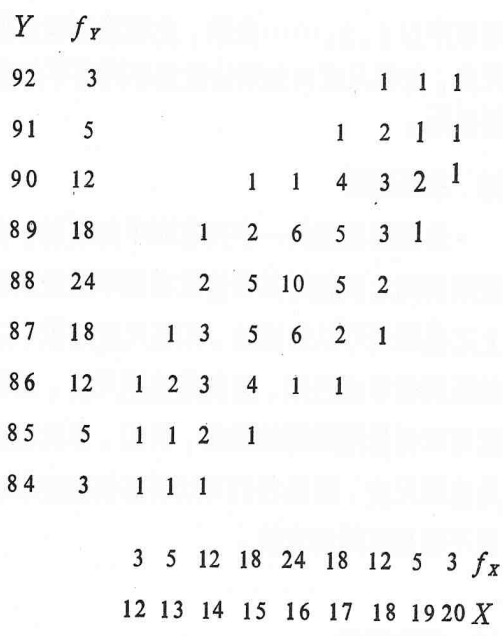


圖 3-2 兩變數間的關係

若我們固定X值，我們可以看出在Y值之間的所有不同值之差異。同樣的，給定Y值，也可看出所有X值之間的差異。例如，令$X=20$，則Y值有三個不同的值，爲 90，91 及 92。又例如，令$Y=86$，則X值有 12，13，14，15，16 及 17六個不同值。

我們將X值排序，給定任一已知行，且允許Y變動，及將Y值排序，給定任一已知列，且允許X變動。因此，在每一行中，排除變數X之效應我們得到Y變數的分配，在每一列中，排除變數Y之效應我們得到X變數的分配。變數效應排除的意思是不考慮影響變數差異的因素，例如，X變數在某一行不影響Y，因爲所有個體在同一行中有同一X值。除了這些X值之外，在這些行中有些因素會影響Y的差異。

3.6 一變數固定變異之效應

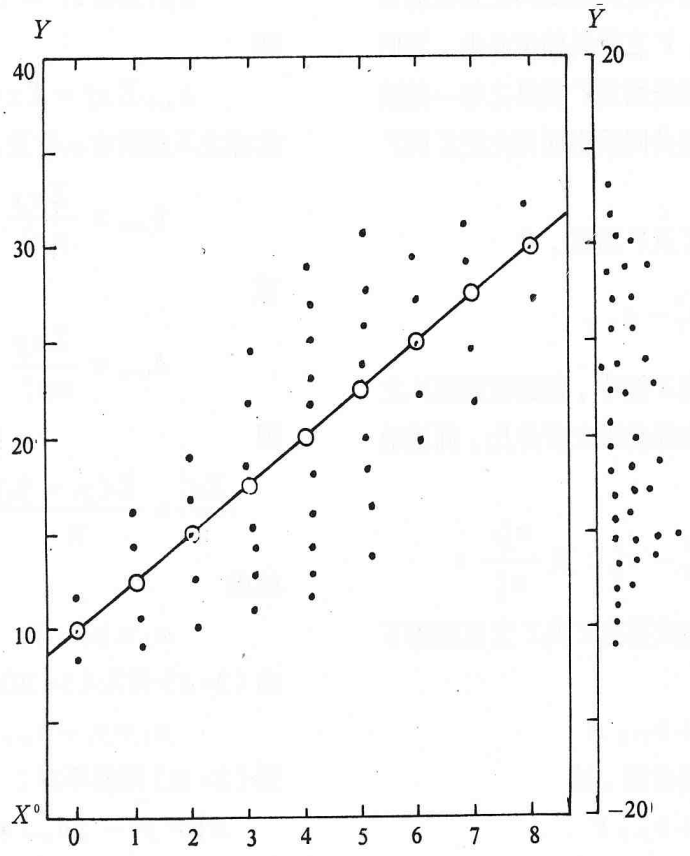
在圖 3-3中之散佈圖表示50 個體在兩變數X及Y上之結合性，每一對用黑點表示。爲方便計，我們考慮Y的變異以表示X與Y關係的結果。因此我們將處理行中之散佈圖。

檢查散佈圖的目的在提供一些資訊。第一，我們可以看到兩變數之數值可以用線性表示來描述。第二，我們可以看到兩變數爲正相關

。一變數之低數值傾向於與另一變數之低數值結合，而一變數之高數值傾向於與另一變數之高數值結合。50個Y值之邊際分配表示於圖3-3之右邊。

假如我們考慮任意行之情形，當X爲常數時我們有Y之分配。在行之變異數可以與Y之總變異比較，其總變異不提到任何X之資訊。此種比較將告訴我們當X爲常數時是否對Y之差異有任何影響。如果有相同X值而Y值有點差異，則我們說它們之間有關係。

在圖3-3，σ_{Yi}值爲Y之標準差，這些值由1.63到4.21，這些值都可以跟Y之總變異



n_i	3	4	5	8	10	8	5	4	3	$\sigma^2_{Y\cdot X}=12.04$
$\bar{Y}_i$	10.0	12.5	15.0	17.5	20.0	22.5	25.0	27.5	30.0	$\sigma_{Y\cdot X}=3.47$
σ^2_{Yi}	2.67	7.25	6.80	17.75	17.60	17.75	6.80	7.25	2.67	$\bar{Y}=20.0$
σ_{Yi}	1.63	2.69	2.61	4.21	4.20	4.21	2.61	2.69	1.63	$\sigma^2_Y=40.04$
										$\sigma_Y=6.33$

圖 3-3　以直線上的點表示所有與該點同一行的值之線性關係

註：$\sigma^2_{Y\cdot X}=\dfrac{n_1\sigma^2_{Y1}+\cdots+n_k\sigma^2_{Yk}}{n}$，$n=n_1+n_2+\cdots+n_k$

及σ^2_Y表示右邊Y邊際分配之變異數，$\sigma^2_Y=\dfrac{1}{n}\Sigma(Y-\bar{Y})^2$

$\sigma_Y = 6.33$ 比較。因此我們可以看出來，當我們固定X值時，Y之變異比Y之總變異小，以此做爲X與Y關係的解釋。

如果X與Y完全無關，在每一行中Y的變異將會跟Y的總變異一樣，即每一行的標準差都是σ_Y。另一方面，若X與Y爲完全相關，則在每一行中Y均無變異，即每一行中之標準差均爲0.00。

在每一行中之標準差與總變異之比較將告訴我們當X固定時，Y之變異縮減多少。我們關心Y的變異之縮減是因爲Y變異之每一縮減會引起更高百分比的共同因素而因此使X與Y更相似。

對於一般變數X及Y而言，令

$$\sigma_{Y'}^2 = \sigma_Y^2 - \sigma_{Y\cdot X}^2$$

因此$\sigma_{Y'}^2/\sigma_Y^2$爲固定X值時，個體在變數Y之變異由X所引起之全部差異之百分比。同樣地，令

$$\sigma_{X'}^2 = \sigma_X^2 - \sigma_{X\cdot Y}^2 \text{ 及 } \frac{\sigma_{X'}^2}{\sigma_X^2} \text{。}$$

假設以下列方程式描述X與Y之直線如下：

$$Y = a_Y + b_{Y\cdot X}\, X$$

此處a_Y與$b_{Y\cdot X}$均爲常數。或

$$X = a_X + b_{X\cdot Y}\, Y$$

此處a_X與$b_{X\cdot Y}$均爲常數。

我們來看看圖3-3中所表示之關係。令$\overline{Y_i}$表示在X_i行中之Y_i值的平均值，此直線方程式

$$\overline{Y_i} = a_Y + b_{Y\cdot X}\, X_i \qquad (3\text{-}1)$$

或

$$\overline{X_i} = a_X + b_{X\cdot Y}\, Y_i \qquad (3\text{-}2)$$

在(3、1)及(3、2)中並不一定表示同一直線。我們若分別將X_i及Y_i變換爲$X_i-\overline{X}$及$Y_i-\overline{Y}$，即將坐標原點移至$(\overline{X},\overline{Y})$，此處$\overline{X}$及$\overline{Y}$分別表示所有$X$及$Y$值之算術平均數，並令 $\overline{x_i} = \frac{1}{n_i}\Sigma x_i$，$x_i = X_i - \overline{X}$ 及 $\overline{y_i} = \frac{1}{n_i}\Sigma y_i$，$y_i = Y_i - \overline{Y}$，則$a_y = 0$或$a_x = 0$，且

$$\overline{y}_i = b_{y\cdot x}\, x_i \qquad (3\text{-}3)$$

或

$$\overline{x}_i = b_{x\cdot y}\, y_i \qquad (3\text{-}4)$$

由(3-3)

$$n_i b_{y\cdot x}\, x_i = \Sigma y_i \qquad (3\text{-}5)$$

將(3-5)兩邊乘x_i，

$$n_i b_{y\cdot x}\, x_i^2 = x_i \Sigma y_i \qquad (3\text{-}6)$$

故

$$b_{y\cdot x}\, \Sigma n_i x_i^2 = (\Sigma x_i)(\Sigma y_i)$$

即

$$b_{y\cdot x}\, \Sigma x_i^2 = \Sigma xy$$

此處之Σ表所有x_i值及$x_i y_i$值之和。因此

$$b_{y\cdot x} = \frac{\Sigma xy}{n\sigma_x^2} \qquad (3\text{-}7)$$

或

$$b_{x\cdot y} = \frac{\Sigma xy}{n\sigma_y^2} \qquad (3\text{-}8)$$

因

$$\frac{\Sigma v_i^2}{n} = \frac{\Sigma (y_i - \overline{y}_i)^2}{n} = \sigma_{y\cdot x}^2 \qquad (3\text{-}9)$$

此處

$$v_i = y_i - \overline{y}_i \qquad (3\text{-}10)$$

由(3-3)代入(3-10)

$$v_i = y_i - b_{y\cdot x}\, x_i \qquad (3\text{-}11)$$

將(3-11)兩邊平方：

$$v_i^2 = y_i^2 - 2b_{y\cdot x}\, x_i y_i + b_{y\cdot x}^2\, x_i^2$$

故

$$\Sigma v_i^2 = \Sigma y_i^2 - 2b_{y\cdot x}\Sigma x_i y_i + b_{y\cdot x}^2\, \Sigma x_i^2$$

兩邊除n，並由$b_{y\cdot x} = \Sigma xy/n\sigma_x^2$得

$$\frac{1}{n}\Sigma v_i^2 = \sigma_y^2 - 2b_{y\cdot x}^2\, \sigma_x^2 + b_{y\cdot x}^2\, \sigma_x^2$$
$$= \sigma_y^2 - b_{y\cdot x}^2\, \sigma_x^2 \qquad (3\text{-}12)$$

由(3-9)及(3-12)，

$$\sigma_{y\cdot x}^2 = \sigma_y^2 - b_{y\cdot x}^2\, \sigma_x^2$$

故

$$\sigma_{y'}^2 = b_{y\cdot x}^2\, \sigma_x^2 \qquad (3\text{-}13)$$

或

$$\sigma_{x'}^2 = b_{x\cdot y}^2\, \sigma_y^2 \qquad (3\text{-}14)$$

由(3-13)及(3-14)

$$\frac{\sigma_{y'}^2}{\sigma_y^2} = b_{y\cdot x}^2 \frac{\sigma_x^2}{\sigma_y^2} \qquad (2\text{-}15)$$

及

$$\frac{\sigma_{x'}^2}{\sigma_x^2} = b_{x\cdot y}^2 \frac{\sigma_y^2}{\sigma_x^2} \qquad (2\text{-}16)$$

3.7 相關係數

若我們分別將 x 及 y 均標準化時，我們可由(3-7)及(3-8)看出 $b_{y\cdot x}$ 及 $b_{x\cdot y}$ 均相同，此時我們不必擔心 x 及 y 所使用的單位。我們可將 x 及 y 之標準分數之相關係數規定爲

$$r_{x\cdot y} = \frac{\Sigma z_x z_y}{n}$$

此處 z_x 及 z_y 分別爲 X 及 Y 之標準分數，即

$$z_x = \frac{X-\overline{X}}{\sigma_X} \text{ 及 } z_y = \frac{Y-\overline{Y}}{\sigma_Y}$$

因此可改寫爲

$$r_{X\cdot Y} = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{n\sigma_X \sigma_Y} \qquad (3\text{-}17)$$

此公式爲皮爾遜係數（Pearsonian Coefficient）。我們可改寫爲

$$r_{X\cdot Y} = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}}} \qquad (3\text{-}18)$$

因此我們可以看出

$$r_{X\cdot Y} = \frac{\sigma_{Y'}}{\sigma_Y}$$

或

$$r_{X\cdot Y} = \frac{\sigma_{X'}}{\sigma_X}$$

我們可以說皮爾遜係數爲解釋變異的百分比。

另外，有兩種相關型態必須考慮。第一種是一個分類變數及另一個爲連續變數；另一種是兩變數均爲分類變數。第一種相關型態之係數稱爲點雙序列係數（point biserial coefficient），以 r_{pbis} 記之。第二種相關型態稱爲 *phi* 係數，以 ϕ 記之。（參考Ghiselli等(1981)）。

令 X 表分類變數及 Y 爲連續變數。此時，X 之值爲 0 或 1。點雙序列係數爲

$$r_{pbis} = \frac{\Sigma Y_1 - \frac{n_1 \Sigma Y}{n}}{\sqrt{n - \frac{n_1^2}{n}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}}} \qquad (3\text{-}19)$$

此處 ΣY_1 表示 $X=1$ 時，所有 Y 值之和，ΣY 爲所有 Y 值和，n_1 表示所有 $X=1$ 之 X 的個數，n 表示所有 X 的個數（參考 Edwards(1976)）。因爲

$$\Sigma XY = \Sigma Y_1 \quad , \quad \Sigma X = \Sigma X^2 = n_1$$

$$\Sigma(X-\overline{X})^2 = n_1 - \frac{n_1^2}{n}$$

代入(3-18)得(3-19)。

我們考慮兩個分類變數 X 及 Y 之相關，此時 X 及 Y 之值均爲 0 或 1。對於 $X=1$ 之個數爲

$$\Sigma X = n_1$$

及

$$\overline{X} = \frac{n_1}{n} = p_x$$

此處 n 爲總個數。同樣地，

$$\Sigma Y = n_2$$

及

$$\overline{Y} = \frac{n_2}{n} = p_y \ ,$$

而對於 $X=1$ 及 $Y=1$ 之個數爲 n_{12}，及 $n_{12}/n = p_{xy}$。因此

$$\Sigma(X-\overline{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$= \Sigma X - \frac{(\Sigma X)^2}{n}$$

$$= n_1 - \frac{n_1^2}{n}$$

同理，

$$\Sigma(Y-\overline{Y})^2 = n_2 - \frac{n_2^2}{n}$$

由(3-18)，

$$r_{X\cdot Y}=\frac{\Sigma XY-\frac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\Sigma X^2-\frac{(\Sigma X)^2}{n}}\sqrt{\Sigma Y^2-\frac{(\Sigma Y)^2}{n}}}$$

$$=\frac{n_{12}-\frac{n_1 n_2}{n}}{\sqrt{n_1-\frac{n_1^2}{n}}\sqrt{n_2-\frac{n_2^2}{n}}}$$

$$=\frac{p_{xy}-p_x p_y}{\sqrt{p_x-p_x^2}\sqrt{p_y-p_y^2}}$$

$$=\frac{p_{xy}-p_x p_y}{\sqrt{p_x q_x}\sqrt{p_y q_y}} \qquad (3\text{-}20)$$

此處$q_x=1-p_x$ 及 $q_y=1-p_y$ 。

公式(3-20)稱爲*phi*係數，以ϕ表示$r_{X\cdot Y}$。

現在將(3-20)改寫其他型式，如下圖

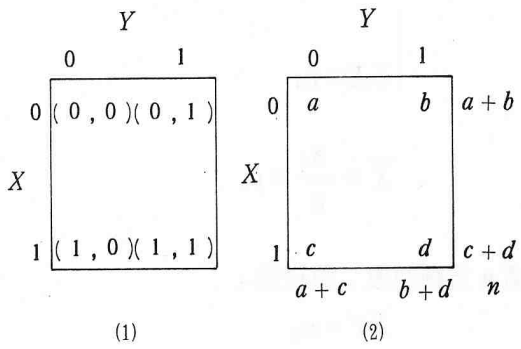


圖 3-4 (1)表示所有可能(X,Y)值之組合
(2)在全部n個值中X及Y之各種值之個數

由圖3-4中，$\Sigma X=\Sigma X^2=c+d$

$\Sigma Y=\Sigma Y^2=b+d$

$\Sigma XY=d$

$n=a+b+c+d$

代入(3-18)，

$$r_{XY}=\frac{d-\frac{(c+d)(b+d)}{n}}{\sqrt{(c+d)-\frac{(c+d)^2}{n}}\sqrt{(b+d)-\frac{(b+d)^2}{n}}}$$

$$=\frac{ad-bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}} \qquad (3\text{-}21)$$

當X及Y均爲有序變數時，都將它們的順序由1至n排序，它們的值均爲由1至n，則

$$\Sigma X=\Sigma Y=\frac{n(n+1)}{2}$$

及

$$\Sigma(X-\overline{X})^2=\Sigma(Y-\overline{Y})^2=\frac{n^3-n}{12}$$

令$D=X-Y$，則$\Sigma D=\Sigma X-\Sigma Y$，因此

$$\overline{D}=\overline{X}-\overline{Y}$$

故

$$\Sigma(D-\overline{D})^2=\Sigma[(X-\overline{X})-(Y-\overline{Y})]^2$$
$$=\Sigma(X-\overline{X})^2+\Sigma(Y-\overline{Y})^2$$
$$-2\Sigma(X-\overline{X})(Y-\overline{Y})$$
$$=\Sigma(X-\overline{X})^2+\Sigma(Y-\overline{Y})^2-$$
$$2r_{XY}\sqrt{\Sigma(X-\overline{X})^2}\sqrt{\Sigma(Y-\overline{Y})^2}$$

即

$$r_{X\cdot Y}=\frac{\Sigma(X-\overline{X})^2+\Sigma(Y-Y)^2-\Sigma(D-\overline{D})^2}{2\sqrt{\Sigma(X-\overline{X})^2}\sqrt{\Sigma(Y-\overline{Y})^2}}$$

簡化得

$$r_{X\cdot Y}=1-\frac{6\Sigma(D-\overline{D})^2}{n^3-n}$$

$$=1-\frac{6\Sigma D^2}{n^3-n} \qquad (3\text{-}21)$$

因爲$\overline{D}=\overline{X}-\overline{Y}=0$，公式(3-21)稱爲有序相關係數(rank order coorrelation coefficient)。

參考資料

1. Edwards, A.L. (1976). *An Introduction to Linear Regression and Correlation*. W.H. Freeman and Company, San Francisco.
2. Ghiselli, E.E., Campbell, J.P. and Zedeck, Z. (1981). *Measurement Theory for the Behavioral Sciences*. W.H. Freeman and Company, San Francisco.

—本文作者任教於師大數學系—