$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2 \sum(y-\overline{y})^2}}$$

相關係數

相關係數正是描述直線關聯性有多強的指標，它可以描述二變數之間直線關聯的方向和強度，通常以符號 r 表示。

以下說明相關係數 r 的原理：

假設我們有 n 個個體的 x 變數(身高)和 y 變數(體重)的資料，每個個體都有配對的 x 和 y，亦即第一位的身高為 x_1，體重為 y_1；第一位的身高為 x_2，體重為 y_2
；第 i 位的身高為 x_i，體重為 y_i。

1. 先求出第一個變數(即身高)的平均數 $\bar{x}$ 及標準差 s_x ，然後算出每個 x 觀測值的標準計分，即計算 $(x_i - \bar{x})/\, s_x$。

2. 再求出第二個變數(即體重)的平均數 $\bar{y}$ 及標準差 s_y ，然後算出每個 y 觀測值的標準計分，即計算 $(y_i - \bar{y})/\, s_y$。

3. 相關係數 r 就是 n 個個體標準計分乘積的平均。

   即： r = $\{[(x_1 - \bar{x})/\, s_x][(y_1 - \bar{y})/\, s_y] + [(x_2 - \bar{x})/\, s_x][(y_2 - \bar{y})/\, s_y] + \cdots + [(x_n - \bar{x})/\, s_x][(y_n - \bar{y})/\, s_y]\}/(n-1)$

4. 若沒有標準化，相關係數亦可由以下公式計算：

$$= \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\sum \overline{x^2} - (\overline{x})^2} \cdot \sqrt{\sum \overline{y^2} - (\overline{y})^2}}$$

相關係數計算起來相當麻煩，還好現在的計算機軟體功能強勁，我們只要輸入數據，可以立刻就得到結果。

比計算相關係數 r 值更重要的是，我們要了解相關係數的涵意如下：

- 相關係數 r 值就是顯示變數之間相關的方向和強弱

  正值代表正相關，負值代表負相關，所以正負是代表方向，而數值大小則代表相關性的強弱，愈接近|1|代表愈強。

- 相關係數 r 值永遠在 -1 ~ +1 之間 (證明就留給統計專家)

  0 代表變數之間完全不相關，愈靠近 0 代表變數間呈愈弱的直線相關性，當 r 愈接近 -1 或 +1 時，代表愈強的直線相關，也就是點的分佈會很接近一條直線。亦即 -1 表示相反的完全直線關係， +1 表示完全的正向直線關係。

- 相關係數 r 不受度量單位的影響

  因為相關係數 r 是用標準計分計算的，不管原來是什麼單位，標準化時分子分母的單位抵銷了，所以本身只是 -1 ~ +1 之間的數值，沒有單位。

- 相關係數與變數的性質無關

  我們有時會設定 x 為解釋變數，y 為反應變數，如果把 x 與 y 的名稱互換，算出來的相關係數還是一樣。

- 相關係數度量的只是二變數間的直線相關強度

  相關係數不能度量變數之間的曲線相關關係，不管它們有多強。

- 相關係數會受到少數離群值的嚴重影響

  當散佈圖明顯出現離群點時， r 值很容易被扭曲。

- 相關係數使用的限制

  對於名目尺度的變數來說，相關係數並無意義。
  列出相關係數的同時，也應該列出各變數的平均數和標準

差，因為它們之間是互有關聯的。

- 相關程度的判定

  相關度的高低一般分為三級：

$\lvert r \rvert < 0.4$	為低度相關
$0.4 \leq \lvert r \rvert < 0.7$	為顯著相關
$0.7 \leq \lvert r \rvert < 1$	為高度相關

- 相關程度不是與 r 值成正比例

  相關係數只是表示關係密切與否的指標而已，我們不可將之視為等距或等比變數。因此，我們不可以說相關係數 r 從 0.4 增加到 0.5 等於相關係數 r 從 0.7 增加到 0.8。當然也不可以說 r = 0.9 是 r = 0.45 的二倍。然而， r = 0.5 卻與 r = - 0.5 表示具有相同程度的密切關係，只是一正一負而已。

- 有相關關係存在不一定有因果關係存在

  這是一個非常重要的觀念，一不小心很容易把相關關係誤認為因果關係。舉個明顯的例子：小學生的腳板大小雖然與數學能力有很明顯的相關，但是不代表二者是因果關係。因為腳板大並不能產生較高的數學能力，而是由於年齡增長自然會在校數學課上得愈多，同時腳板也長得愈大，才會呈現正向的相關關係。所以當我們在解釋相關關係時要特別小心，除非有充分證據，否則不可輕易斷言有因果關係。