THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
Exercises on Correlation

Learning Outcomes 學習目標

- Understand the importance and limitations of correlation analysis,
  理解相關分析的重要性與限制，
- Distinguish between:
  區分以下情況：
  (a) Linear and non-linear correlation,
  線性與非線性相關，
  (b) Positive and negative correlation,
  正相關與負相關，
- Calculate and interpret the correlation coefficient for:
  計算並解釋相關係數，適用於：
  (i) Individual (ungrouped) observations,
  個別（未分組）觀察值，
  (ii) Bivariate grouped data.
  雙變量分組資料。

Karl Pearson's Method 卡爾皮爾森方法

Karl Pearson's method of calculating the coefficient of correlation is based on the covariance of two variables in a series. In fact, there are two methods to calculate the coefficient of correlation:
卡爾皮爾森（Karl Pearson）計算**相關係數**的方法是基於一個數列中兩個變數的**共變異數**。事實上，計算相關係數有兩種方法：

1. Direct method 直接法
2. Short-cut method 捷徑法

Direct Method 直接法

In this method, the coefficient of correlation (denoted by r) is calculated as the ratio of the covariance of the two variables to the product of their standard deviations.
在此方法中，相關係數（記為r）定義為兩個變數的共變異數與其標準差乘積之比。

Symbolically:
符號表示如下：

$$r = \frac{\sum xy}{N\sigma_x\sigma_y}$$

where:
其中：

- $x = (X - \bar{X})$, $y = (Y - \bar{Y})$ (deviations of X and Y from their arithmetic means),
  $x = (X - \bar{X})$，$y = (Y - \bar{Y})$：X 與Y 對其算術平均數的**偏差**，
- $\sum xy$ = sum of the product of deviations in X and Y,
  $\sum xy$：X 與Y 偏差乘積的總和，
- σ_x = standard deviation of X, σ_y = standard deviation of Y,
  σ_x：X 的**標準差**，σ_y：Y 的**標準差**，
- N = total number of pairs of observations.
  N：觀測值配對的總數。

A simplified version of this formula (derived below) is:
此公式的簡化版本（推導如下）為：

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \times \sqrt{\sum y^2}}$$

Derivation of the Simplified Formula 簡化公式的推導

Recall that:
試回想

$$\sigma_x = \sqrt{\frac{\sum x^2}{N}}, \quad \sigma_y = \sqrt{\frac{\sum y^2}{N}}$$

Substitute into the original formula:
代入原式：

$$r = \frac{\sum xy}{N \cdot \sqrt{\frac{\sum x^2}{N}} \cdot \sqrt{\frac{\sum y^2}{N}}} = \frac{\sum xy}{N \cdot \frac{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}}{N}} = \frac{\sum xy}{\sqrt{\sum x^2} \times \sqrt{\sum y^2}}$$

This formula applies only when deviations are taken from the actual arithmetic means of the two series. It is also called the *covariance method* or *product-moment method*.
此公式僅適用於偏差取自兩個數列的實際算術平均數。它也被稱為共變異數法或積差法。

Steps to Calculate r (Direct Method)
計算r 的步驟（直接法）

To calculate the coefficient of correlation:
為了計算**相關係數**：

1. Compute the arithmetic means of X ($\bar{X}$) and Y ($\bar{Y}$).
   計算X 的算術平均數（$\bar{X}$）與Y 的算術平均數（$\bar{Y}$）。
2. Calculate deviations: $x = X - \bar{X}$ and $y = Y - \bar{Y}$.
   計算偏差：$x = X - \bar{X}$，$y = Y - \bar{Y}$。
3. Square the deviations: compute $\sum x^2$ (sum of squared x-deviations) and $\sum y^2$ (sum of squared y-deviations).
   將偏差平方：計算$\sum x^2$（x 偏差平方和）與$\sum y^2$（y 偏差平方和）。

4. Compute $\sum xy$: multiply corresponding x and y values, then sum the products.
計算$\sum xy$：將對應的x 與y 相乘，然後求乘積總和。

5. Substitute $\sum xy$, $\sum x^2$, and $\sum y^2$ into the simplified formula for r.
將$\sum xy$、$\sum x^2$ 和$\sum y^2$ 代入簡化公式以求得r。

Example 1: Calculate the Coefficient of Correlation
例題1：計算相關係數

For the data:
就該組數據：

X	2	3	4	5	6
Y	7	9	10	14	15

Solution 解答

First, set up a table to compute the required values (where $x = X - \bar{X}$ and $y = Y - \bar{Y}$):
首先，建立表格計算所需數值（其中$x = X - \bar{X}$，$y = Y - \bar{Y}$）：

X	Y	$x = X - \bar{X}$	x^2	$y = Y - \bar{Y}$	y^2	xy
2	7	$2-4=-2$	$(-2)^2=4$	$7-11=-4$	$(-4)^2=16$	$(-2)(-4)=8$
3	9	$3-4=-1$	$(-1)^2=1$	$9-11=-2$	$(-2)^2=4$	$(-1)(-2)=2$
4	10	$4-4=0$	$0^2=0$	$10-11=-1$	$(-1)^2=1$	$(0)(-1)=0$
5	14	$5-4=1$	$1^2=1$	$14-11=3$	$3^2=9$	$(1)(3)=3$
6	15	$6-4=2$	$2^2=4$	$15-11=4$	$4^2=16$	$(2)(4)=8$
$\sum X=20$	$\sum Y=55$		$\sum x^2=10$		$\sum y^2=46$	$\sum xy=21$

Next, compute the means (where $N = 5$ is the number of observations):
接著，計算平均數（其中$N = 5$ 為觀測值數量）：

$$\bar{X} = \frac{\sum X}{N} = \_\_\_\_\_\_, \quad \bar{Y} = \frac{\sum Y}{N} = \_\_\_\_\_\_$$

Substitute into the Pearson correlation coefficient formula:
代入皮爾森相關係數公式：

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \times \sqrt{\sum y^2}}$$

Calculate the square roots (rounded to two decimal places):
計算平方根（四捨五入至小數點後兩位）：

$$r = \_\_\_\_\_\_$$

The value $r \approx$ ______ indicates an almost perfect positive correlation between X and Y.
$r \approx$ ______ 的數值表示X 與Y 之間接近完全正相關。

Note on the Range of r
關於r 取值範圍的說明

The coefficient of correlation r always lies between -1 and $+1$:
相關係數r 的取值範圍始終介於-1 與$+1$ 之間：

- $r = 0$: No correlation between the series.
  兩個數列之間**無相關**。
- $r = -1$: Perfect negative correlation.
  完全負相關。
- $r = +1$: Perfect positive correlation.
  完全正相關。

Calculating Coefficient of Correlation by Short-cut Method
以捷徑法計算相關係數

The following steps are involved in calculating the coefficient of correlation using this method:
使用此方法計算**相關係數**的步驟如下：

1. Choose *convenient values* as the *assumed means* of the two series X and Y.
   選擇方便的數值作為兩個數列X 與Y 的假設平均數。
2. Deviations (dx and dy, instead of x and y) are obtained from the assumed means (in the same manner as deviations from actual means).
   以假設平均數計算偏差（使用dx 與dy，而非x 與y），其計算方式與實際平均數的偏差相同。
3. Compute the sum of the dx and dy columns: $\sum dx$ and $\sum dy$.
   計算dx 與dy 欄的總和：$\sum dx$ 與$\sum dy$。
4. Square the deviations dx and dy, then calculate their totals: $\sum dx^2$ and $\sum dy^2$.
   將dx 與dy 平方，並計算其總和：$\sum dx^2$ 與$\sum dy^2$。
5. Compute $\sum dxdy$, the sum of the products of deviations taken from the assumed means of the two series.
   計算$\sum dxdy$，即兩數列相對於假設平均數的偏差乘積總和。

Example 2
例題2

A company manufactures different types of electrical appliances. It uses radio for advertising its products. The table below shows six weeks of radio time (X, in minutes) and the number of electrical appliances sold (Y) over the last six weeks:
一家公司生產不同類型的電器產品，並使用電台廣告。下表顯示過去六週的電台廣告時間（X，單位：分鐘）與電器銷售數量（Y）：

X	25	18	32	21	35	29
Y	16	11	20	15	26	28

Calculate the coefficient of correlation between X and Y.
計算X 與Y 之間的相關係數。

Solution 解答

First, set up a table to compute the required values (note: $A.M.$ = Assumed Mean):
首先，建立表格計算所需數值（注意：$A.M.$ 表示假設平均數）：

X Series			Y Series			
X	$dx = X - A.M._X$	dx^2	Y	$dy = Y - A.M._Y$	dy^2	$dx \cdot dy$
	$(A.M._X = 25)$			$(A.M._Y = 20)$		
25	$25 - 25 = 0$	$0^2 = 0$	16	$16 - 20 = -4$	$(-4)^2 = 16$	$0 \cdot (-4) = 0$
18	$18 - 25 = -7$	$(-7)^2 = 49$	11	$11 - 20 = -9$	$(-9)^2 = 81$	$(-7) \cdot (-9) = 63$
32	$32 - 25 = 7$	$7^2 = 49$	20	$20 - 20 = 0$	$0^2 = 0$	$7 \cdot 0 = 0$
21	$21 - 25 = -4$	$(-4)^2 = 16$	15	$15 - 20 = -5$	$(-5)^2 = 25$	$(-4) \cdot (-5) = 20$
35	$35 - 25 = 10$	$10^2 = 100$	26	$26 - 20 = 6$	$6^2 = 36$	$10 \cdot 6 = 60$
29	$29 - 25 = 4$	$4^2 = 16$	28	$28 - 20 = 8$	$8^2 = 64$	$4 \cdot 8 = 32$
$\sum X = 160$	$\sum dx = 10$	$\sum dx^2 = 230$	$\sum Y = 116$	$\sum dy = -4$	$\sum dy^2 = 222$	$\sum dxdy = 175$

The formula for r (Short-cut Method) is:
捷徑法計算r 的公式為：

$$r = \frac{\sum dxdy - \frac{(\sum dx)(\sum dy)}{N}}{\sqrt{\sum dx^2 - \frac{(\sum dx)^2}{N}} \times \sqrt{\sum dy^2 - \frac{(\sum dy)^2}{N}}}$$

where $N = 6$ (number of observations).
其中$N = 6$（觀測值數量）。

1: Compute the Numerator 計算分子

$$\text{Numerator} = \sum dxdy - \frac{(\sum dx)(\sum dy)}{N} = \_\_\_\_\_$$

2: Compute the Denominator (First Term) 計算分母（第一項）

$$\sqrt{\sum dx^2 - \frac{(\sum dx)^2}{N}} = \_\_\_\_\_$$

3: Compute the Denominator (Second Term) 計算分母（第二項）

$$\sqrt{\sum dy^2 - \frac{(\sum dy)^2}{N}} = \_\_\_\_\_$$

4: Calculate r 計算r

$$r = \_\_\_\_\_$$

This value ($r \approx$ ______) indicates a high degree of positive correlation between radio time (X) and the number of appliances sold (Y).
此值（$r \approx$ ______）表示電台廣告時間（X）與電器銷售數量（Y）之間存在高度正相關。

Note on the Short-cut Formula 捷徑公式說明

When comparing the short-cut formula to the direct method (using actual means), $\frac{(\sum dx)(\sum dy)}{N}$ is the *correction factor* for the numerator, and $\frac{(\sum dx)^2}{N}$, $\frac{(\sum dy)^2}{N}$ are correction factors for the denominator. These account for using *assumed means* instead of actual arithmetic means.
將捷徑公式與直接法（使用實際平均數）比較時，$\frac{(\sum dx)(\sum dy)}{N}$ 是分子的校正因子，而$\frac{(\sum dx)^2}{N}$ 和$\frac{(\sum dy)^2}{N}$ 是分母的校正因子。這些校正因子用於處理使用假設平均數而非實際算術平均數的情況。

Alternative Method (Using Original Values)
替代方法（使用原始數值）

There is another method to calculate the coefficient of correlation: instead of taking deviations from an assumed mean, we can compute r directly using the original values of the two series. The formula for this method is:
另一種計算相關係數的方法：不從假設平均數計算偏差，而是直接使用兩個數列的原始數值計算r。此方法的公式為：

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

Example 3
例題3

We use this formula with the *same data* as Example 2 (radio time X and appliances sold Y) to calculate the Pearson correlation coefficient using the alternative (direct) method:
我們使用與例題2相同的數據（電台廣告時間X 與電器銷售數量Y），以替代方法（直接法）計算皮爾森相關係數：

Solution 解答

First, set up a table to compute the required sums ($N = 6$, number of observations):
首先，建立表格計算所需的總和（$N = 6$，觀測值數量）：

X	X^2	Y	Y^2	XY
25	$25^2 = 625$	16	$16^2 = 256$	$25 \times 16 = 400$
18	$18^2 = 324$	11	$11^2 = 121$	$18 \times 11 = 198$
32	$32^2 = 1024$	20	$20^2 = 400$	$32 \times 20 = 640$
21	$21^2 = 441$	15	$15^2 = 225$	$21 \times 15 = 315$
35	$35^2 = 1225$	26	$26^2 = 676$	$35 \times 26 = 910$
29	$29^2 = 841$	28	$28^2 = 784$	$29 \times 28 = 812$
$\sum X = 160$	$\sum X^2 = 4480$	$\sum Y = 116$	$\sum Y^2 = 2462$	$\sum XY = 3275$

We now compute r step-by-step using the alternative Pearson correlation formula:
我們現在逐步使用替代的皮爾森相關係數公式計算r：

1: Calculate the Numerator 計算分子

$$\text{Numerator} = \sum XY - \frac{(\sum X)(\sum Y)}{N}$$

Substitute the values:
代入數值：

$$\text{Numerator} = \_\_\_\_\_\_$$

2: Calculate the First Term in the Denominator 計算分母的第一項

$$\sum X^2 - \frac{(\sum X)^2}{N} = \_\_\_\_\_\_$$

3: Calculate the Second Term in the Denominator 計算分母的第二項

$$\sum Y^2 - \frac{(\sum Y)^2}{N} = \_\_\_\_\_\_$$

4: Calculate the Denominator 計算分母

$$\text{Denominator} = \_\_\_\_\_\_$$

5: Compute r 計算r

$$r = \_\_\_\_\_\_$$

This matches the result from the Short-cut Method ($r \approx$ ______), confirming that the Alternative Method (using original values) yields the same coefficient of correlation—indicating a high positive correlation between radio time (X) and the number of appliances sold (Y).
此結果與捷徑法計算的結果（$r \approx$ ______）一致，證實替代方法（使用原始數值）得到相同的相關係數，表示電台廣告時間（X）與電器銷售數量（Y）之間存在高度正相關。

Correlation Coefficient for Grouped Data 分組資料的相關係數

When calculating the correlation coefficient between two variables from **grouped data**, we use a two-way frequency table (rows = one variable, columns = the other variable). The table cells contain frequencies (number of observations) for each combination of class intervals. We denote:
當從**分組資料**計算兩個變數之間的相關係數時，我們使用雙向次數表（列= 一個變數，行= 另一個變數）。表格中的儲存格包含每個組距組合的次數（觀測值數量）。記號如下：

- m = number of rows (excluding totals),
- m = 列數（不含總計），

- n = number of columns (excluding totals),
- n = 行數（不含總計），

- f = frequency in a cell,
- f = 儲存格中的次數，

- N = total frequency ($\sum f$).
- N = 總次數（$\sum f$）。

Formulas for Correlation Coefficient (r) with Grouped Data
分組資料相關係數的公式

The formula (using deviations from assumed means) is:
公式（使用假設平均數的偏差）為：

$$r = \frac{\frac{\sum f d_x d_y}{N} - \frac{(\sum f d_x)(\sum f d_y)}{N^2}}{\sqrt{\left(\frac{\sum f d_x^2}{N} - \frac{(\sum f d_x)^2}{N^2}\right)\left(\frac{\sum f d_y^2}{N} - \frac{(\sum f d_y)^2}{N^2}\right)}}$$

A simplified version (equivalent) is:
簡化版本（等價）為：

$$r = \frac{N \sum f d_x d_y - (\sum f d_x)(\sum f d_y)}{\sqrt{\left[N \sum f d_x^2 - (\sum f d_x)^2\right]\left[N \sum f d_y^2 - (\sum f d_y)^2\right]}}$$

Where:
其中：

- d_x = deviation of X (row variable) from its assumed mean,
- $d_x = X$（列變數）對其假設平均數的偏差，

- d_y = deviation of Y (column variable) from its assumed mean,
- $d_y = Y$（行變數）對其假設平均數的偏差，

- $f d_x d_y$ = product of d_x, d_y, and cell frequency,
- $f d_x d_y = d_x$、d_y 與儲存格次數的乘積，

- $f d_x^2 = d_x^2 \times$ cell frequency,
- $f d_x^2 = d_x^2 \times$ 儲存格次數，

- $f d_y^2 = d_y^2 \times$ cell frequency.
- $f d_y^2 = d_y^2 \times$ 儲存格次數。

Steps to Calculate Correlation for Grouped Data
計算分組資料相關的步驟

i. Record mid-points of class intervals for X (columns) and Y (rows).
記錄X（行）與Y（列）的組距中點。

ii. Choose an assumed mean for X (M_V) and Y (M'_V), then calculate deviations ($d_x = X - M_V$, $d_y = Y - M'_V$).
選擇X 與Y 的假設平均數（M_V 與M'_V），然後計算偏差（$d_x = X - M_V$，$d_y = Y - M'_V$）。

iii. (Optional) Simplify calculations by dividing deviations by a common factor (step-deviations).
（可選）將偏差除以一個公因數（步級偏差）以簡化計算。

iv. Compute $f d_x$, $f d_y$, $f d_x d_y$, $f d_x^2$, $f d_y^2$ for each cell.
計算每個儲存格的$f d_x$、$f d_y$、$f d_x d_y$、$f d_x^2$、$f d_y^2$。

v. Sum these values to get $\sum f d_x$, $\sum f d_y$, $\sum f d_x d_y$, $\sum f d_x^2$, $\sum f d_y^2$.
將這些值加總，得到$\sum f d_x$、$\sum f d_y$、$\sum f d_x d_y$、$\sum f d_x^2$、$\sum f d_y^2$。

vi. Substitute into the correlation formula.
代入相關公式。

Example 4: Sales Revenue vs Advertising Expenditure
例題4：銷售收入vs 廣告支出

Table 1 shows grouped data for Sales Revenue (Y, HKD) and Advertising Expenditure (X, HKD). Calculate the Pearson correlation coefficient for the grouped data.
表1 顯示銷售收入（Y，港幣）與廣告支出（X，港幣）的分組資料。計算分組資料的皮爾森相關係數。

Table 1: Two-Way Frequency Table (Sales Revenue vs Advertising Expenditure)
雙向次數表（銷售收入vs 廣告支出）

Sales Revenue (HKD) (Y)	Advertising Expenditure (HKD) (X)				Total
	5–15	15–25	25–35	35–45	
75–125	3	4	4	8	19
125–175	8	6	5	7	26
175–225	2	2	3	4	11
225–275	3	3	2	2	10
Total	16	15	14	21	66

1: Define Mid-Points and Assumed Means 定義中點與假設平均數

- X (Advertising Expenditure) mid-points: 10 (5–15), 20 (15–25), 30 (25–35), 40 (35–45).
Assumed mean for X (M_V) = 20 (mid-point of 15–25).
Deviation for X (scaled by class width $w_x = 10$): $d_x = \frac{X-20}{10}$; d_x values: $\frac{10-20}{10} = -1$, $\frac{20-20}{10} = 0$, $\frac{30-20}{10} = 1$, $\frac{40-20}{10} = 2$.
X (廣告開支) 中值：10 (5–15), 20 (15–25), 30 (25–35), 40 (35–45)。
X 的假定平均數(M_V) =20 (15–25 區間的中值)。
X 的偏差（按組距$w_x = 10$ 縮放）：$d_x = \frac{X-20}{10}$; d_x 的值為：$\frac{10-20}{10} = -1$, $\frac{20-20}{10} = 0$, $\frac{30-20}{10} = 1$, $\frac{40-20}{10} = 2$.

- Y (Sales Revenue) mid-points: 100 (75–125), 150 (125–175), 200 (175–225), 250 (225–275).
Assumed mean for Y (M'_V) = 150 (mid-point of 125–175).
Deviation for Y (scaled by class width $w_y = 50$): $d_y = \frac{Y-150}{50}$; d_y values: $\frac{100-150}{50} = -1$, $\frac{150-150}{50} = 0$, $\frac{200-150}{50} = 1$, $\frac{250-150}{50} = 2$.
Y (銷售收入) 中值：100 (75–125), 150 (125–175), 200 (175–225), 250 (225–275)。
Y 的假定平均數(M'_V) =150 (25–175 區間的中值)。
Y 的偏差（按組距$w_y = 50$ 縮放）：$d_y = \frac{Y-150}{50}$; d_y 的值為：$\frac{100-150}{50} = -1$, $\frac{150-150}{50} = 0$, $\frac{200-150}{50} = 1$, $\frac{250-150}{50} = 2$.

Table 2: Computation of Correlation Coefficient (Example 4) 相關係數計算（例題4）

Y $M'_V = 150$ $d_y \downarrow$	X (Advertising Expenditure) $M_V = 20$ $d_x \rightarrow$				Total (f)	fd_y	fd_y^2	$\sum fd_xd_y$
	5–15 ($d_x = -1$)	15–25 ($d_x = 0$)	25–35 ($d_x = 1$)	35–45 ($d_x = 2$)				
75–125 ($d_y = -1$)	$3(-1)(-1) = 3$	$4(0)(-1) = 0$	$4(1)(-1) = -4$	$8(2)(-1) = -16$	19	$19(-1) = -19$	$19(1) = 19$	$3 + 0 - 4 - 16 = -17$
125–175 ($d_y = 0$)	$8(-1)(0) = 0$	$6(0)(0) = 0$	$5(1)(0) = 0$	$7(2)(0) = 0$	26	$26(0) = 0$	$26(0) = 0$	$0 + 0 + 0 + 0 = 0$
175–225 ($d_y = 1$)	$2(-1)(1) = -2$	$2(0)(1) = 0$	$3(1)(1) = 3$	$4(2)(1) = 8$	11	$11(1) = 11$	$11(1) = 11$	$-2 + 0 + 3 + 8 = 9$
225–275 ($d_y = 2$)	$3(-1)(2) = -6$	$3(0)(2) = 0$	$2(1)(2) = 4$	$2(2)(2) = 8$	10	$10(2) = 20$	$10(4) = 40$	$-6 + 0 + 4 + 8 = 6$
Total (f)	16	15	14	21	$N = 66$	$\sum fd_y = 12$	$\sum fd_y^2 = 70$	$\sum fd_xd_y = -2$
fd_x	$16(-1) = -16$	$15(0) = 0$	$14(1) = 14$	$21(2) = 42$	$\sum fd_x = 40$			
fd_x^2	$16(1) = 16$	$15(0) = 0$	$14(1) = 14$	$21(4) = 84$	$\sum fd_x^2 = 114$			
$fdxdy$	-5	0	4	0	$\sum fdxdy = -2$			

2: Compute Correlation Coefficient 計算相關係數

We use the simplified formula for grouped data (Pearson correlation coefficient):
我們使用分組資料的簡化公式（皮爾森相關係數）：

$$r = \frac{N \sum fd_xd_y - (\sum fd_x)(\sum fd_y)}{\sqrt{[N \sum fd_x^2 - (\sum fd_x)^2] \times [N \sum fd_y^2 - (\sum fd_y)^2]}}$$

Substitute the values from Table 2:
代入表2 的數值：

- $N = 66$, $\sum fd_xd_y = -2$, $\sum fd_x = 40$, $\sum fd_y = 12$,

- $\sum fd_x^2 = 114$, $\sum fd_y^2 = 70$.

Step-by-Step Arithmetic:
逐步計算：

1. Numerator:
   分子：

   ______

2. Denominator (first part: X term):
   分母（第一部分：X 項）：

   ______

3. Denominator (second part: Y term):
   分母（第二部分：Y 項）：

   ______

4. Combine denominator terms (exact product + square root):
   合併分母各項（準確積值+ 平方根）：

   ______

5. Final r (full precision):
   最終r 值（全精確度）：

   $r =$ ______

Alternative Calculation (Original Formula):
替代計算（原始公式）：

$$r = \frac{\frac{\sum f d_x d_y}{N} - \frac{(\sum f d_x)(\sum f d_y)}{N^2}}{\sqrt{\left(\frac{\sum f d_x^2}{N} - \frac{(\sum f d_x)^2}{N^2}\right) \times \left(\frac{\sum f d_y^2}{N} - \frac{(\sum f d_y)^2}{N^2}\right)}}$$

- Numerator:
  分子：

$$\_\_\_\_\_\_$$

- X term:
  X 項：

$$\_\_\_\_\_\_$$

- Y term:
  Y 項：

$$\_\_\_\_\_\_$$

- Denominator:
  分母：

$$\_\_\_\_\_\_$$

- Final r (full precision):
  最終r 值（全精確度）：

$$r = \_\_\_\_\_\_$$

(matches simplified formula, confirming consistency).
（與簡化公式一致，確認計算正確。）

Example 5: Family Income vs Food Expenditure (Engel's Law)
例題5：家庭收入vs 食品支出（恩格爾定律）

Tables 3–6 show grouped data for Family Income (X, HKD '000) and Food Expenditure (Y, %) for 100 families. Calculate the Pearson correlation coefficient and interpret the result (Engel's Law: as income rises, the share of expenditure on food falls).
表3–6 顯示100 個家庭的家庭收入（X，單位：港幣千元）與食品支出（Y，百分比）的分組資料。計算皮爾森相關係數並解釋結果（恩格爾定律：收入增加時，食品支出佔比下降）。

Table 3: Family Income vs Food Expenditure (100 Families)
家庭收入vs 食品支出（100 個家庭）

Food Expenditure (%) (Y)	Family Income (HKD ’000) (X)				Total
	5–10	10–15	15–20	20–25	
10–15	–	–	–	3	3
15–20	–	4	9	4	17
20–25	7	6	12	5	30
25–30	3	10	19	8	40
Total	10	20	40	30	100

1: Define Mid-Points and Step-Deviations 定義中點與步級偏差

- X (Family Income) mid-points: 7.5 (5–10), 12.5 (10–15), 17.5 (15–20), 22.5 (20–25). Assumed mean for $X = 17.5$ (mid-point of 15–20), common factor (class width) = 5. Step-deviation for X: $u_i = \frac{X-17.5}{5}$: u_i values: $\frac{7.5-17.5}{5} = -2$, $\frac{12.5-17.5}{5} = -1$, $\frac{17.5-17.5}{5} = 0$, $\frac{22.5-17.5}{5} = 1$.
X（家庭收入）中點：7.5（5–10）、12.5（10–15）、17.5（15–20）、22.5（20–25）。假設平均數= 17.5（15–20 的中點），公因數（組距）= 5。X 的步級偏差：$u_i = \frac{X-17.5}{5}$：u_i 值：$-2, -1, 0, 1$。

- Y (Food Expenditure) mid-points: 12.5 (10–15), 17.5 (15–20), 22.5 (20–25), 27.5 (25–30). Assumed mean for $Y = 22.5$ (mid-point of 20–25), common factor (class width) = 5. Step-deviation for Y: $v_i = \frac{Y-22.5}{5}$: v_i values: $\frac{12.5-22.5}{5} = -2$, $\frac{17.5-22.5}{5} = -1$, $\frac{22.5-22.5}{5} = 0$, $\frac{27.5-22.5}{5} = 1$.
Y（食品支出）中點：12.5（10–15）、17.5（15–20）、22.5（20–25）、27.5（25–30）。假設平均數= 22.5（20–25 的中點），公因數（組距）= 5。Y 的步級偏差：$v_i = \frac{Y-22.5}{5}$：v_i 值：$-2, -1, 0, 1$。

2: Calculation Tables 計算表格

Table 4: Calculation of $\sum fu_i$ and $\sum fu_i^2$ (Family Income, X)
計算$\sum fu_i$ 與$\sum fu_i^2$（家庭收入，X）

u_i	f_i (Total Frequency)	f_iu_i	$f_iu_i^2$
-2	10	$10(-2) = -20$	$10(4) = 40$
-1	20	$20(-1) = -20$	$20(1) = 20$
0	40	$40(0) = 0$	$40(0) = 0$
1	30	$30(1) = 30$	$30(1) = 30$
Total	100	$\sum fu_i = -10$	$\sum fu_i^2 = 90$

Table 5: Calculation of $\sum fv_i$ and $\sum fv_i^2$ (Food Expenditure, Y)
計算$\sum fv_i$ 與$\sum fv_i^2$（食品支出，Y）

v_i	f_i (Total Frequency)	f_iv_i	$f_iv_i^2$
-2	3	$3(-2) = -6$	$3(4) = 12$
-1	17	$17(-1) = -17$	$17(1) = 17$
0	30	$30(0) = 0$	$30(0) = 0$
1	40	$40(1) = 40$	$40(1) = 40$
Total	100	$\sum fv_i = 17$	$\sum fv_i^2 = 69$

Table 6: Calculation of $\sum fu_iv_i$ (Cross Product) 計算$\sum fu_iv_i$（交叉乘積）

$v_i \downarrow$	$u_i \rightarrow$	-2 (5–10)	-1 (10–15)	0 (15–20)	1 (20–25)	Total (fv_i)
-2 (10–15)		–	–	–	$3(-2)(1) = -6$	$3(-2) = -6$
-1 (15–20)		–	$4(-1)(-1) = 4$	$9(-1)(0) = 0$	$4(-1)(1) = -4$	$17(-1) = -17$
0 (20–25)		$7(0)(-2) = 0$	$6(0)(-1) = 0$	$12(0)(0) = 0$	$5(0)(1) = 0$	$30(0) = 0$
1 (25–30)		$3(1)(-2) = -6$	$10(1)(-1) = -10$	$19(1)(0) = 0$	$8(1)(1) = 8$	$40(1) = 40$
$\sum fu_iv_i$ (column-wise)		-6	-6	0	-2	$\sum fu_iv_i = -14$

3: Compute Correlation Coefficient 計算相關係數

Use the step-deviation formula for grouped data (equivalent to Pearson's correlation coefficient):
使用分組資料的步級偏差公式（等價於皮爾森相關係數）：

$$r = \frac{N\sum fuv - (\sum fu)(\sum fv)}{\sqrt{[N\sum fu^2 - (\sum fu)^2] \times [N\sum fv^2 - (\sum fv)^2]}}$$

Substitute corrected values from Tables 4–6:
代入表4–6 的數值：

- $N = 100$, $\sum fuv = -14$, $\sum fu = -10$, $\sum fv = 17$,
- $\sum fu^2 = 90$, $\sum fv^2 = 69$.

Step-by-Step Arithmetic (Full Precision):
逐步計算（完整精度）：

1. Numerator:
   分子：

   ______

2. Denominator (first part: u term):
   分母（第一部分：u 項）：

   ______

3. Denominator (second part: v term):
   分母（第二部分：v 項）：

   ______

4. Combine denominator terms:
合併分母各項：

$$\_\_\_\_\_\_$$

5. Final r (unrounded + rounded):
最終r 值（未經捨入+ 經捨入）：

$$r = \_\_\_\_\_\_$$

4: Interpretation (Engel's Law) 解釋（恩格爾定律）

The correlation coefficient $r \approx -0.16$ indicates a weak negative correlation between family income (HKD '000) and the percentage of income spent on food. While this aligns with Engel's Law (higher income $\rightarrow$ lower food expenditure share), the weak correlation suggests other factors (e.g., household size, food preferences) also influence food expenditure patterns for this sample of 100 families.
相關係數$r \approx -0.16$ 表示家庭收入（單位：港幣千元）與收入中用於食品支出的百分比之間存在弱負相關。此結果與恩格爾定律一致（收入愈高$\rightarrow$ 食品支出占比愈低），但由於相關性較弱，顯示其他因素（例如：家庭規模、飲食偏好）亦會影響本樣本中100 個家庭的食品支出模式。

t Test for a Correlation Coefficient 相關係數的t 檢定

In the examples given earlier in this chapter, the correlation coefficient r was computed to measure the strength of the relationship between two variables. Note that we often use the sample correlation coefficient for descriptive purposes as a point estimate of the population correlation coefficient ρ. This means r is treated as an estimate of the parameter ρ, but r can only be validly used as an estimate of ρ if the assumption of bivariate normality (normal distribution for both variables) holds.
在本章前面的例子中，我們計算了相關係數r 來衡量兩個變數之間的關係強度。注意，我們通常將樣本相關係數用作母體相關係數ρ 的點估計。這意味著r 被視為參數ρ 的估計值，但只有在雙變量常態性假設（兩個變數皆呈常態分佈）成立時，r 才能有效地用作ρ 的估計。

The most frequently used test to examine whether two variables X and Y are linearly correlated is the t-test. To apply this test, we first specify the null and alternative hypotheses as follows:
檢驗兩個變數X 與Y 是否線性相關的最常用方法是t 檢定。要應用此檢定，我們首先設定虛無假設與對立假設如下：

- $H_0 : \rho = 0$ (No linear correlation between X and Y)
  $H_0 : \rho = 0$（X 與Y 之間無線性相關）
- $H_1 : \rho \neq 0$ (A linear correlation exists between X and Y)
  $H_1 : \rho \neq 0$（X 與Y 之間存在線性相關）

where ρ (rho) denotes the population correlation coefficient.
其中ρ 表示總體相關係數。

The formula for the t-test statistic is:
t 檢定統計量的公式為：

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

where:
其中：

- r = sample correlation coefficient,
  r = 樣本相關係數，

- ρ = hypothesized population correlation coefficient (typically 0 for this test),
- ρ = 假設的總體相關係數（此檢定通常為0），

- n = sample size,
  n = 樣本大小，

- $n - 2$ = degrees of freedom (df) for the t-distribution.
  $n - 2 = t$ 分佈的自由度（df）。

The test statistic t follows a t-distribution with $n - 2$ degrees of freedom. We illustrate this with an example.
檢定統計量t 遵循自由度為$n - 2$ 的t 分佈。以下用一個例子來說明。

Example 6
例題6

Refer to Example 3, where the sample correlation coefficient $r = 0.84$ and sample size $n = 6$. Test whether there is a statistically significant linear association between advertising expenditure and sales revenue at the 5% significance level.
參考例題3，其中樣本相關係數$r = 0.84$，樣本大小$n = 6$。在5% 顯著水準下，檢定廣告支出與銷售收入之間是否存在統計上顯著的線性關聯。

Solution 解答

We apply the t-test using the formula above, with $\rho = 0$ (null hypothesis of no correlation):
我們使用上述公式進行t 檢定，其中$\rho = 0$（虛無假設為無相關）：

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

1: Substitute values 代入數值

$r = 0.84$, $\rho = 0$, $n = 6$:

$$t = \frac{0.84 - 0}{\sqrt{\frac{1-(0.84)^2}{6-2}}}$$

2: Calculate the denominator (full precision) 計算分母（完整精度）

1. Compute the squared correlation: $(0.84)^2 = 0.7056$
   計算相關係數平方：$(0.84)^2 = 0.7056$

2. Compute $1 - r^2$: $1 - 0.7056 = 0.2944$
   計算$1 - r^2$：$1 - 0.7056 = 0.2944$
3. Divide by degrees of freedom ($n - 2 = 4$): $\frac{0.2944}{4} = 0.0736$
   除以自由度($n - 2 = 4$)：$\frac{0.2944}{4} = 0.0736$
4. Take the square root: $\sqrt{0.0736} = 0.2712932 \approx 0.2713$
   取平方根：$\sqrt{0.0736} = 0.2712932 \approx 0.2713$

3: Compute the t-statistic 計算t 統計量

$$t = \frac{0.84}{0.2712932} = 3.09627 \approx 3.10$$

4: Compare to critical value 與臨界值比較

For $df = 4$ and a two-tailed 5% significance level, the critical t-value is 2.776.
對於自由度$df = 4$ 和雙尾5% 顯著水準，臨界t 值為2.776。

Since the calculated $t = 3.10 > 2.776$ (critical value), we reject the null hypothesis $H_0 : \rho = 0$.
由於計算得到的$t = 3.10 > 2.776$（臨界值），我們拒絕虛無假設$H_0 : \rho = 0$。

Conclusion 結論

There is a statistically significant linear association between advertising expenditure and sales revenue at the 5% significance level.
在5% 顯著水準下，廣告支出與銷售收入之間存在統計上顯著的線性關聯。

RANK CORRELATION 等級相關

Sometimes we encounter statistical series that are *ranked by size* (rather than measured with exact magnitudes, which cannot be ascertained). In such cases, Karl Pearson's coefficient of correlation is not suitable. Instead, Spearman's Rank Correlation (or simply rank correlation) is used. This method relies on the *ranks (or order)* of observations (not their specific values) and involves simple calculations.
有時我們會遇到按大小排序的統計數列（而非以精確數值測量，因為無法確定其精確量值）。在這種情況下，卡爾皮爾森的相關係數並不適用，而是採用**史皮爾曼等級相關**（Spearman's Rank Correlation，簡稱**等級相關**）。此方法依賴觀測值的等級（或順序），而非其具體數值，且計算相對簡單。

Types of Rank Correlation Problems 等級相關問題的類型

We encounter two types of problems in rank correlation:
在等級相關中，我們會遇到兩種類型的問題：

1. When *ranks are already given* for the series.
   當等級已經給定於數列中。

2. When *ranks are not given* (we assign ranks to raw data first).
   當等級尚未給定（需先對原始資料進行排序並分配等級）。

Case 1: Ranks Are Given
情況一：等級已給定

When ranks are provided:
當等級已經提供時：

1. Compute the *difference* between the two ranks for each observation: $d = R_1 - R_2$ (where R_1, R_2 are ranks for the two series).
   計算每個觀測值在兩個數列中的等級差：$d = R_1 - R_2$（其中R_1, R_2 為兩個數列的等級）。

2. Square these differences and calculate their total: $\sum d^2$.
   將這些差值平方並計算其總和：$\sum d^2$。

3. Apply Spearman's formula:
   套用史皮爾曼公式：

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

   where:
   其中：

   - r_s = Spearman's rank correlation coefficient,
     r_s：史皮爾曼等級相關係數，
   - N = number of observations.
     N：觀測值的數量。

Case 2: Ranks Are Not Given
情況二：未給定名次

When raw data (not ranks) are provided:
當提供的是原始資料（而非名次）時：

1. Assign ranks to each series (either rank the *highest value as 1* or the *lowest value as 1*—use the same rule consistently for both series).
   為每一組資料分配名次（可以選擇將最大值排名為*1*，或將最小值排名為*1*——兩組資料必須一致採用相同的規則）。

2. For *tied ranks* (identical observations): Assign the *average rank* to all tied values. For example, if two observations tie for 7th and 8th place, assign $\frac{7+8}{2} = 7.5$ to both.
   對於並列名次（相同觀測值）：將平均名次分配給所有並列值。例如，若兩個觀測值並列第7 與第8 名，則分配$\frac{7+8}{2} = 7.5$ 給兩者。

3. Proceed with steps 1–3 from Case 1 (compute $d, \sum d^2$, then apply the formula).
   繼續執行情況一的步驟1–3（計算$d, \sum d^2$，然後套用公式）。

Example 7: Spearman's Rank Correlation (Training vs Sales Performance)
例題7：史皮爾曼等級相關（培訓表現vs 銷售表現）

10 salesmen at a company completed a month of training, took a test (ranked by performance), and were later ranked on their sales performance (6 months post-training). Their ranks are shown below:
某公司的10 名銷售員完成了一個月的培訓，並參加了測試（按表現排名），之後在培訓後六個月根據銷售表現進行排名。他們的排名如下：

Salesmen	1	2	3	4	5	6	7	8	9	10
Ranks (Training)	4	6	1	3	9	7	10	2	8	5
Ranks (Sales Performance)	5	8	3	1	7	6	9	2	10	4

Calculate Spearman's coefficient of rank correlation and comment on the result.
計算史皮爾曼等級相關係數並解釋結果。

Solution 解答

First, compute the rank difference d = Training Rank − Sales Rank and squared difference d^2 for each salesman (Table 7):
首先，計算每個銷售員的等級差d = 培訓排名 − 銷售排名 與平方差d^2（表7）：

Table 7: Calculation of Rank Differences (d) and Squared Differences (d^2)
等級差與平方差計算

Salesmen	Training Rank (R_1)	Sales Rank (R_2)	Difference ($d = R_1 - R_2$)	Squared Difference (d^2)
1	4	5	$4 - 5 = -1$	$(-1)^2 = 1$
2	6	8	$6 - 8 = -2$	$(-2)^2 = 4$
3	1	3	$1 - 3 = -2$	$(-2)^2 = 4$
4	3	1	$3 - 1 = 2$	$2^2 = 4$
5	9	7	$9 - 7 = 2$	$2^2 = 4$
6	7	6	$7 - 6 = 1$	$1^2 = 1$
7	10	9	$10 - 9 = 1$	$1^2 = 1$
8	2	2	$2 - 2 = 0$	$0^2 = 0$
9	8	10	$8 - 10 = -2$	$(-2)^2 = 4$
10	5	4	$5 - 4 = 1$	$1^2 = 1$
				$\sum d^2 = 24$

Apply Spearman's Rank Correlation Formula
套用史皮爾曼等級相關公式

Spearman's rank correlation coefficient formula (for no tied ranks) is:
史皮爾曼等級相關係數公式（無並列名次時）為：

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

where:
其中：

- $N = 10$ (number of salesmen),
 $N = 10$（銷售員人數），

- $\sum d^2 = 24$ (sum of squared rank differences).
 $\sum d^2 = 24$（等級差平方和）。

Substitute values (full precision):
代入數值（完整精度）：

$$r_s = 1 - \frac{6(24)}{10(10^2 - 1)} = \_\_\_\_\_\_$$

Interpretation 解釋

A rank correlation coefficient of $r_s =$ ______ indicates a very high degree of positive correlation between performance in training and subsequent sales performance. This means the two rankings are strongly associated: salesmen who performed well in training tend to achieve high sales ranks, and vice versa.
等級相關係數$r_s =$ ______ 表示培訓表現與後續銷售表現之間存在非常高的正相關。這意味著兩個排名之間有强烈的關聯：在培訓中表現良好的銷售員往往獲得較高的銷售排名，反之亦然。

Significance of r_s 等級相關係數的顯著性

For small samples ($N \leq 30$), the distribution of r_s is non-normal, so the t-test is inappropriate. Instead, use the Spearman's Rank Correlation Significance Test:
對於小樣本（$N \leq 30$），r_s 的分佈非常態，因此t 檢定不適用。應改用史皮爾曼等級相關顯著性檢定：

- *Null Hypothesis (H_0)*: $r_s = 0$ (no linear correlation between training ranks and sales ranks).
 虛無假設（H_0）：$r_s = 0$（培訓排名與銷售排名之間無線性相關）。
- *Alternative Hypothesis (H_1)*: $r_s \neq 0$ (a linear correlation exists between the two sets of ranks).
 對立假設（H_1）：$r_s \neq 0$（兩組排名之間存在線性相關）。
- For $N = 10$ and a 5% (two-tailed) significance level, the critical value of r_s is 0.6364.
 對於$N = 10$ 和5%（雙尾）顯著水準，r_s 的臨界值為0.6364。

Since the calculated $r_s = 0.855 > 0.6364$ (critical value), we reject the null hypothesis (H_0). This confirms that training performance and sales performance for the 10 salesmen are statistically significantly associated at the 5% level.
由於計算得到的$r_s = 0.855 > 0.6364$（臨界值），我們拒絕虛無假設（H_0）。這證實了在5%顯著水準下，這10 名銷售員的培訓表現與銷售表現之間存在統計上顯著的關聯。

Example 8: Spearman's Rank Correlation (with Tied Ranks)
例題8：史皮爾曼等級相關（含並列名次）

Calculate Spearman's coefficient of rank correlation between two assessments (Internal and External) of workers' performance (higher scores = higher rank, with rank 1 assigned to the highest score):
計算工人表現的兩項評核（內部評核與外部評核）之間的史皮爾曼等級相關係數（分數越高排名越高，排名1 分配給最高分數）：

Name of Workers	A	B	C	D	E	F	G	H	I
Internal Assessment (Int. Asmt.)	51	63	73	46	50	60	47	36	60
External Assessment (Ext. Asmt.)	49	72	74	44	58	66	50	30	35

Solution 解答

First, assign ranks to each assessment (note: tied scores require *average ranks*). Ranks are assigned such that higher marks receive a lower numerical rank (rank 1 = highest score, consistent for both series):
首先，為每項評核分配等級（注意：相同分數需要平均排名）。分配等級時，較高分數獲得較小的數字排名（排名1 = 最高分數，兩組系列保持一致）：

Name	Int. Asmt.	Ranks R_1	Ext. Asmt.	Ranks R_2	Difference ($d = R_1 - R_2$)	Squared Difference (d^2)
A	51	5	49	6	$5 - 6 = -1$	1.00
B	63	2	72	2	$2 - 2 = 0$	0.00
C	73	1	74	1	$1 - 1 = 0$	0.00
D	46	8	44	7	$8 - 7 = 1$	1.00
E	50	6	58	4	$6 - 4 = 2$	4.00
F	60	3.5	66	3	$3.5 - 3 = 0.5$	0.25
G	47	7	50	5	$7 - 5 = 2$	4.00
H	36	9	30	9	$9 - 9 = 0$	0.00
I	60	3.5	35	8	$3.5 - 8 = -4.5$	20.25
						$\sum d^2 = 30.50$

Tied Ranks Explanation (Internal Assessment):
並列名次說明（內部評核）：

Workers F and I both scored 60 in the Internal Assessment. Without ties, their ranks would be 3 and 4 (since 60 is the 3rd/4th highest score). The average rank for tied scores is calculated as:
工人F 和I 在內部評核中都得到了60 分。若無並列，他們的排名將為3 和4（因為60 是第3/4 高分）。並列分數的平均排名計算如下：

$$\text{Average Rank for F/I} = \frac{3+4}{2} = 3.5$$

No ties exist in the External Assessment, so all R_2 ranks are unique.
外部評核中沒有並列分數，因此所有R_2 排名都是唯一的。

Apply Spearman's Rank Correlation Formula (for Tied Ranks) 套用史皮爾曼等級相關公式（含並列名次）

Spearman's rank correlation formula (adjustment for tied ranks is not required here, as ties are only in one series and the basic formula is sufficient for this example):
史皮爾曼等級相關公式（此例中並列名次僅出現在一個數列，基本公式已足夠）：

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

where:
其中：

- $N = 9$ (number of workers),
  $N = 9$（工人人數），

- $\sum d^2 = 30.50$ (sum of squared rank differences).
  $\sum d^2 = 30.50$（等級差平方和）。

Substitute values (full precision):
代入數值（完整精度）：

$$r_s = 1 - \frac{6(30.50)}{9(9^2-1)} = 1 - \frac{183}{9(80)} = 1 - \frac{183}{720} = 1 - 0.2541667 = 0.7458333 \approx 0.746 \approx 0.75$$

Significance Test for r_s 等級相關係數的顯著性檢定

For a two-tailed test at the 5% significance level with $N = 9$:
對於$N = 9$ 在5%（雙尾）顯著水準下：

- *Null Hypothesis (H_0)*: $r_s = 0$ (no linear correlation between Internal and External assessment ranks).
  虛無假設（H_0）：$r_s = 0$（內部評核排名與外部評核排名之間無線性相關）。

- *Alternative Hypothesis (H_1)*: $r_s \neq 0$ (a linear correlation exists between the two sets of ranks).
  對立假設（H_1）：$r_s \neq 0$（兩組排名之間存在線性相關）。

- Critical value of $r_s = 0.6833$ (from Spearman's rank correlation critical value tables).
  r_s 的臨界值為0.6833（參照史皮爾曼等級相關臨界值表）。

Since the calculated $r_s = 0.746 > 0.6833$ (critical value), we reject the null hypothesis (H_0). This confirms a statistically significant high positive correlation between Internal and External assessment ranks for the 9 workers.
由於計算得到的$r_s = 0.746 > 0.6833$（臨界值），我們拒絕虛無假設（H_0）。這證實了這9 名工人的內部評核排名與外部評核排名之間存在統計上顯著的高度正相關。

Example 9: Correcting a Rank Correlation Error 例題9：修正等級相關錯誤

The rank correlation coefficient between 10 students' marks in Statistics and Accountancy was initially calculated as 0.2. Later, it was discovered that the rank difference for one student was

incorrectly recorded as 9 (instead of the correct value of 7). Calculate the corrected coefficient of rank correlation.
10 名學生在統計學與會計學的成績等級相關係數最初計算為0.2。後來發現，其中一名學生的等級差被錯誤記錄為9（正確值應為7）。計算修正後的等級相關係數。

Solution 解答

We start with Spearman's rank correlation formula to solve for the *original* sum of squared rank differences ($\sum d^2$):
我們從史皮爾曼等級相關公式開始，解出原始的等級差平方和（$\sum d^2$）：

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

Given:
已知

- Initial $r_s = 0.2$,
  初始$r_s = 0.2$，
- $N = 10$ (number of students).
  $N = 10$（學生人數）。

Rearrange the formula to isolate $\sum d^2$:
重新排列公式以解出$\sum d^2$：

$$0.2 = 1 - \frac{6\sum d^2}{10(10^2-1)}$$
$$\frac{6\sum d^2}{10(99)} = 1 - 0.2 = 0.8$$
$$\frac{6\sum d^2}{990} = 0.8$$
$$\sum d^2 = \frac{0.8 \times 990}{6} = 132$$

Correct $\sum d^2$
修正後的$\sum d^2$

The rank difference for one student was incorrectly recorded as 9 (squared difference $= 9^2 = 81$) instead of 7 (correct squared difference $= 7^2 = 49$). Adjust $\sum d^2$ by subtracting the incorrect squared difference and adding the correct one:
一名學生的等級差被錯誤記錄為9（平方差$= 9^2 = 81$），正確值應為7（平方差$= 7^2 = 49$）。修正$\sum d^2$ 時，減去錯誤的平方差並加上正確的平方差：

$$\text{Corrected } \sum d^2 = 132 - 9^2 + 7^2 = 132 - 81 + 49 = 100$$

Compute the Correct r_s
計算正確的r_s

Substitute the corrected $\sum d^2 = 100$ back into Spearman's formula (full precision):
將修正後的$\sum d^2 = 100$ 代回史皮爾曼公式（完整精度）：

$$r_s = 1 - \frac{6(100)}{10(10^2-1)} = 1 - \frac{600}{990} = 1 - 0.6060606 = 0.3939394 \approx 0.394 \approx 0.39$$

The corrected rank correlation coefficient is 0.39 (or 0.394 when rounded to three decimal places).
修正後的等級相關係數為0.39（四捨五入至三位小數為0.394）。

Limitations of Spearman's Method of Correlation 史皮爾曼相關方法的限制

Spearman's rank correlation (r_s) is a *distribution-free (nonparametric)* measure of monotonic correlation. This means no strict assumptions are made about the underlying probability distribution of the population from which sample observations are drawn. However, this flexibility comes with tradeoffs—r_s may be less statistically powerful than Pearson's correlation (a parametric measure) when the data meet the assumptions of normality.
史皮爾曼等級相關（r_s）是一種無分佈（非參數）的單調相關性度量。這意味著我們不需對樣本觀測值所屬的總體機率分佈作嚴格假設。然而，這種靈活性亦有其權衡——當數據符合正態分佈的假設時，r_s 的統計功效（statistical power）可能低於皮爾森相關係數（一種參數度量）。

Key limitations of Spearman's correlation include:
史皮爾曼相關的主要限制包括：

1. While r_s is simpler to compute than Pearson's r, it is not designed for grouped frequency distributions (it requires individual rank data, not aggregated group data).
雖然r_s 的計算比皮爾森相關r 更簡單，但它並不適用於分組次數分配（需要個別名次資料，而非彙總群組資料）。

2. For large samples ($N \geq 30$)—where ranks are not pre-assigned and must be manually calculated from raw data—the process becomes tedious and time-consuming. This is a major practical limitation in applied research.
對於大型樣本（$N \geq 30$）——若名次未事先給定，必須從原始資料手動計算名次，過程將變得繁瑣且耗時。這是應用研究中的一項重大實務限制。

Example 10: Find the Number of Observations 例題10：求觀測值數量

Given:
已知：
$\sum x^2 = 90$, $\sigma_y = 2.5$, $\sum xy = 60$, $r = 0.8$
(where x, y are deviations from their respective means).
其中x, y 分別為對其平均數的偏差）。

We use the formula for Pearson's correlation coefficient (when deviations are used):
我們使用皮爾森相關係數公式（使用偏差時）：

$$r = \frac{\sum xy}{n \cdot \sigma_x \cdot \sigma_y}$$

First, compute σ_x (standard deviation of x-series):

首先，計算σ_x（x 數列的標準差）：

$$\sigma_x = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{90}{n}}$$

1: Substitute Values into the Formula 代入公式

$$0.8 = \frac{60}{n \cdot \sqrt{\frac{90}{n}} \cdot 2.5}$$

Simplify the denominator:
化簡分母：

$$n \cdot \sqrt{\frac{90}{n}} \cdot 2.5 = 2.5 \cdot \sqrt{90n}$$

2: Solve for n 解n

Rearrange the equation:
重排方程：

$$0.8 = \frac{60}{2.5 \cdot \sqrt{90n}} \implies 2.5 \cdot \sqrt{90n} = \frac{60}{0.8} = 75$$

$$\sqrt{90n} = \frac{75}{2.5} = 30$$

Square both sides:
兩邊平方：

$$90n = 30^2 = 900 \implies n = \frac{900}{90} = 10$$

The number of observations is $\boxed{10}$.
觀測值數量為$\boxed{10}$。

Example 11: Calculate Correlation Coefficient 例題11：計算相關係數

Given:
已知：
$n = 12$, $\sum X = 96$, $\sum Y = 120$, $\sum(X-8)^2 = 150$, $\sum(Y-10)^2 = 200$, $\sum(X-8)(Y-10) = 50$.

1: Verify Arithmetic Means 驗證算術平均數

$$\bar{X} = \frac{\sum X}{n} = \frac{96}{12} = 8, \quad \bar{Y} = \frac{\sum Y}{n} = \frac{120}{12} = 10$$

Thus:
因此：

- $\sum(X-8)^2 = \sum x^2 = 150$ (squared deviations of X from $\bar{X}$),
  $\sum(X-8)^2 = \sum x^2 = 150$（$X$ 對$\bar{X}$ 的偏差平方和），

- $\sum(Y-10)^2 = \sum y^2 = 200$ (squared deviations of Y from $\bar{Y}$),
  $\sum(Y-10)^2 = \sum y^2 = 200$（$Y$ 對$\bar{Y}$ 的偏差平方和），

- $\sum(X-8)(Y-10) = \sum xy = 50$ (product of deviations).
  $\sum(X-8)(Y-10) = \sum xy = 50$（偏差乘積）。

2: Apply Pearson's Correlation Formula 應用皮爾森相關公式

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

Substitute values:
代入數值：

$$r = \frac{50}{\sqrt{150 \times 200}} = \frac{50}{\sqrt{30000}} = \frac{50}{173.2} \approx 0.29$$

The correlation coefficient between X and Y is $\boxed{0.29}$ (weak positive correlation).
X 與Y 之間的相關係數為$\boxed{0.29}$（弱正相關）。

Example 12: Correcting the Correlation Coefficient 例題12：修正相關係數

A person calculated the correlation coefficient between X and Y and obtained:
某人計算了X 與Y 的相關係數，得到：
$n = 30$, $\sum X = 120$, $\sum X^2 = 600$, $\sum Y = 90$, $\sum Y^2 = 250$, $\sum XY = 356$.

Later, it was found that two pairs $(8, 10)$ and $(12, 7)$ were incorrectly entered (correct pairs: $(8, 12)$ and $(10, 8)$). We need to adjust the sums and compute the correct correlation coefficient.
後來發現有兩對數據$(8, 10)$ 和$(12, 7)$ 輸入錯誤（正確配對應為$(8, 12)$ 和$(10, 8)$）。我們需要調整總和並計算正確的相關係數。

1: Adjust the Sums 調整總和

To correct the sums, *subtract the incorrect values and add the correct values*:
為修正總和，減去錯誤的值並加上正確的值：

$$\begin{aligned}
\text{Corrected } \sum X &= 120 - 8 - 12 + 8 + 10 = 118, \\
\text{Corrected } \sum X^2 &= 600 - 8^2 - 12^2 + 8^2 + 10^2 = 600 - 64 - 144 + 64 + 100 = 556, \\
\text{Corrected } \sum Y &= 90 - 10 - 7 + 12 + 8 = 93, \\
\text{Corrected } \sum Y^2 &= 250 - 10^2 - 7^2 + 12^2 + 8^2 = 250 - 100 - 49 + 144 + 64 = 309, \\
\text{Corrected } \sum XY &= 356 - (8 \times 10) - (12 \times 7) + (8 \times 12) + (10 \times 8) \\
&= 356 - 80 - 84 + 96 + 80 = 368.
\end{aligned}$$

2: Apply the Correlation Formula 應用相關公式

The formula for Pearson's correlation coefficient (using original values) is:
皮爾森相關係數公式（使用原始數值）為：

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

3: Compute the Numerator 計算分子

$$\text{Numerator} = 368 - \frac{118 \times 93}{30} = 368 - \frac{10974}{30} = 368 - 365.8 = 2.2$$

4: Compute the Denominator Terms 計算分母各項

First term (for X):
第一項（X）：

$$\sum X^2 - \frac{(\sum X)^2}{n} = 556 - \frac{118^2}{30} = 556 - \frac{13924}{30} = 556 - 464.13 = 91.87$$

Second term (for Y):
第二項（Y）：

$$\sum Y^2 - \frac{(\sum Y)^2}{n} = 309 - \frac{93^2}{30} = 309 - \frac{8649}{30} = 309 - 288.3 = 20.7$$

5: Compute the Denominator 計算分母

$$\text{Denominator} = \sqrt{91.87 \times 20.7} = \sqrt{1901.709} \approx 43.61$$

6: Calculate the Correct r 計算正確的r

$$r = \frac{2.2}{43.61} \approx 0.05$$

The correct correlation coefficient is $\boxed{0.05}$ (very weak positive correlation).
正確的相關係數為$\boxed{0.05}$（非常弱的正相關）。

Example 13: Correlation Between Aptitude Score and Productivity Index
例題13：能力傾向分數與生產力指數的相關

Data for 6 workers (Aptitude Score X, Productivity Index Y):
6 名工人的數據（能力傾向分數X，生產力指數Y）：

Workers	A	B	C	D	E	F
Aptitude Score (X)	9	18	18	20	20	23
Productivity Index (Y)	23	23	33	42	29	32

1: Compute Sums (Worksheet) 計算總和（工作表）

Workers	X	Y	X^2	Y^2	XY
A	9	23	$9^2 = 81$	$23^2 = 529$	$9 \times 23 = 207$
B	18	23	$18^2 = 324$	$23^2 = 529$	$18 \times 23 = 414$
C	18	33	$18^2 = 324$	$33^2 = 1089$	$18 \times 33 = 594$
D	20	42	$20^2 = 400$	$42^2 = 1764$	$20 \times 42 = 840$
E	20	29	$20^2 = 400$	$29^2 = 841$	$20 \times 29 = 580$
F	23	32	$23^2 = 529$	$32^2 = 1024$	$23 \times 32 = 736$
	$\sum X = 108$	$\sum Y = 182$	$\sum X^2 = 2058$	$\sum Y^2 = 5776$	$\sum XY = 3371$

2: Apply the Correlation Formula 應用相關公式

Using $n = 6$:
使用$n = 6$：

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

3: Compute the Numerator 計算分子

$$\text{Numerator} = 3371 - \frac{108 \times 182}{6} = 3371 - \frac{19656}{6} = 3371 - 3276 = 95$$

4: Compute the Denominator Terms 計算分母各項

First term (for X):
第一項（X）：

$$\sum X^2 - \frac{(\sum X)^2}{n} = 2058 - \frac{108^2}{6} = 2058 - \frac{11664}{6} = 2058 - 1944 = 114$$

Second term (for Y):
第二項（Y）：

$$\sum Y^2 - \frac{(\sum Y)^2}{n} = 5776 - \frac{182^2}{6} = 5776 - \frac{33124}{6} = 5776 - 5520.67 = 255.33$$

5: Compute the Denominator 計算分母

$$\text{Denominator} = \sqrt{114 \times 255.33} = \sqrt{29107.62} \approx 170.61$$

6: Calculate r 計算r

$$r = \frac{95}{170.61} \approx 0.56$$

The correlation coefficient between Aptitude Score and Productivity Index is $\boxed{0.56}$ (moderate positive correlation).
能力傾向分數與生產力指數之間的相關係數為$\boxed{0.56}$（中等正相關）。

Example 14
例題14

The following are the monthly figures on advertising expenditure and sales of a firm. It is generally found that advertising expenditure has an impact on sales after two months. Allowing for this time lag, calculate the coefficient of correlation between expenditure on advertisement and sales.
以下是一家公司的每月廣告支出與銷售額數據。一般發現廣告支出在兩個月後對銷售額產生影響。考慮此時間滯後，計算廣告支出與銷售額之間的相關係數。

Month	Ad. Expenditure ('000) HKD	Sales ('000 HKD)
January	50	1200
February	60	1500
March	70	1600
April	90	2000
May	120	2200
June	150	2500
July	140	2400
August	160	2600
September	170	2800
October	190	2900
November	200	3100
December	250	3900

Solution 解答

Since advertising expenditure impacts sales after 2 months, we align sales data with advertising expenditure from 2 months prior:
由於廣告支出在兩個月後影響銷售，我們將銷售數據與兩個月前的廣告支出對齊：

- January–October advertising expenditure is paired with March–December sales (10 pairs total).
  一月至十月的廣告支出與三月至十二月的銷售額配對（共10 對）。

To simplify calculations:
為簡化計算：

- Divide each ad expenditure value by 10 (denoted x).
  將每個廣告支出值除以10（記為x）。

- Divide each sales value by 100 (denoted y).
  將每個銷售額值除以100（記為y）。

Table 8: Worksheet for Correlation Calculation 相關計算工作表

Month	Ad Exp. (x)	$dx = x - \bar{x}$	dx^2	Sales (y)	$dy = y - \bar{y}$	dy^2	$dxdy$
January	5	$5 - 12 = -7$	49	16	$16 - 26 = -10$	100	$(-7)(-10) = 70$
February	6	$6 - 12 = -6$	36	20	$20 - 26 = -6$	36	$(-6)(-6) = 36$
March	7	$7 - 12 = -5$	25	22	$22 - 26 = -4$	16	$(-5)(-4) = 20$
April	9	$9 - 12 = -3$	9	25	$25 - 26 = -1$	1	$(-3)(-1) = 3$
May	12	$12 - 12 = 0$	0	24	$24 - 26 = -2$	4	$(0)(-2) = 0$
June	15	$15 - 12 = 3$	9	26	$26 - 26 = 0$	0	$(3)(0) = 0$
July	14	$14 - 12 = 2$	4	28	$28 - 26 = 2$	4	$(2)(2) = 4$
August	16	$16 - 12 = 4$	16	29	$29 - 26 = 3$	9	$(4)(3) = 12$
September	17	$17 - 12 = 5$	25	31	$31 - 26 = 5$	25	$(5)(5) = 25$
October	19	$19 - 12 = 7$	49	39	$39 - 26 = 13$	169	$(7)(13) = 91$
Total	$\sum x = 120$		$\sum dx^2 = 222$	$\sum y = 260$		$\sum dy^2 = 364$	$\sum dxdy = 261$

First, calculate the means:
首先，計算平均數：

$$\bar{x} = \frac{\sum x}{n} = \frac{120}{10} = 12, \quad \bar{y} = \frac{\sum y}{n} = \frac{260}{10} = 26$$

The Pearson correlation coefficient is:
皮爾森相關係數為：

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2 \times \sum dy^2}}$$

Substitute values:
代入數值：

$$r = \frac{261}{\sqrt{222 \times 364}} = \frac{261}{\sqrt{80808}} = \frac{261}{284.27} \approx 0.92$$

Example 15
例題15

The director of a management training programme wants to know whether there is a positive association between a trainee's pre-programme score and their post-programme score. The scores of 10 trainees are:
某管理培訓課程的總監想知道學員在課程前的分數與課程後的分數之間是否存在正相關。10 名學員的分數如下：

Trainee	1	2	3	4	5	6	7	8	9	10
Rank Score 1	1	4	10	8	5	7	3	2	6	9
Rank Score 2	2	3	9	10	5	6	1	4	7	8

Determine the degree of association between pre-training and post-training scores.
確定培訓前與培訓後分數之間的關聯程度。

Solution 解答

We use Spearman's Rank Correlation Coefficient (since data are ranks). The formula is:
我們使用史皮爾曼等級相關係數（由於數據是等級）。公式為：

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

where d = Rank difference = $X - Y$, and $n = 10$.
其中d = 等級差 = $X - Y$，$n = 10$。

Table 9: Worksheet for Rank Differences 等級差計算表

Trainee No.	Score 1 (X)	Score 2 (Y)	Rank diff. $d = X - Y$	$d^2 = (X-Y)^2$
1	1	2	$1 - 2 = -1$	1
2	4	3	$4 - 3 = 1$	1
3	10	9	$10 - 9 = 1$	1
4	8	10	$8 - 10 = -2$	4
5	5	5	$5 - 5 = 0$	0
6	7	6	$7 - 6 = 1$	1
7	3	1	$3 - 1 = 2$	4
8	2	4	$2 - 4 = -2$	4
9	6	7	$6 - 7 = -1$	1
10	9	8	$9 - 8 = 1$	1
Total			$\sum d = 0$	$\sum d^2 = 18$

Substitute into the Spearman formula:
代入史皮爾曼公式：

$$r_s = 1 - \frac{6(18)}{10(10^2-1)} = 1 - \frac{108}{10(99)} = 1 - \frac{108}{990} = 1 - 0.11 = 0.89$$

Significance Test 顯著性檢定

To test if r_s is statistically significant:
檢驗r_s 是否統計上顯著：

- Null Hypothesis: $H_0 : \rho_s = 0$ (no association)
  虛無假設：$H_0 : \rho_s = 0$（無關聯）
- Alternative Hypothesis: $H_1 : \rho_s > 0$ (positive association)
  對立假設：$H_1 : \rho_s > 0$（正相關）

For $n = 10$ and 5% significance (one-tailed), the critical value of r_s is 0.6364.
對於$n = 10$ 和5% 顯著水準（單尾），r_s 的臨界值為0.6364。

Since the calculated $r_s = 0.89 > 0.6364$, we reject H_0. The association between pre-training and post-training scores is statistically significant.
由於計算得到的$r_s = 0.89 > 0.6364$，我們拒絕H_0。培訓前與培訓後分數之間的關聯在統計上是顯著的。

Problem 16: Spearman's Rank Correlation for TV Programme Ratings
問題16：電視節目評級的史皮爾曼等級相關

Two individuals were asked to watch ten specified TV programmes and evaluate them by assigning a rank from 1 (lowest) to 10 (highest). Their rankings are provided in the table below. Calculate Spearman's coefficient of correlation for the two sets of ratings.
兩名人士被要求觀看十個指定的電視節目，並按1（最低）至10（最高）分配等級進行評價。他們的排名如下表所示。計算這兩組評級的史皮爾曼相關係數。

Table 10: Rankings of TV Programmes by Two Individuals 兩名人士對電視節目的排名

TV Programme	Ranks by X	Ranks by Y
A	4	2
B	6	3
C	3	4
D	9	9
E	1	5
F	5	7
G	2	7
H	7	1
I	10	8
J	8	6

1: Recall Spearman's Rank Correlation Formula 回顧史皮爾曼等級相關公式

Spearman's rank correlation coefficient (r_s) measures the strength of association between two ranked variables. The formula for untied ranks is:
史皮爾曼等級相關係數（r_s）衡量兩個等級變數之間的關聯強度。無並列名次時的公式為：

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

Where:
其中：

- $d = X - Y$ (difference between ranks for the same programme),
  $d = X - Y$（同一節目的等級差），

- $\sum d^2$ = sum of squared rank differences,
  $\sum d^2$ = 等級差平方和，

- n = number of paired observations (here, $n = 10$).
  n = 配對觀測值數量（此處$n = 10$）。

2: Calculate Rank Differences (d) and Squared Differences (d^2)
計算等級差（d）與平方差（d^2）

First, create a worksheet to compute d and d^2 for each programme:
首先，建立工作表計算每個節目的d 和d^2：

Table 11: Worksheet for Rank Difference Calculation 等級差計算表

TV Programme	X (Rank)	Y (Rank)	$d = X - Y$	$d^2 = (X - Y)^2$
A	4	2	$4 - 2 = 2$	$2^2 = 4$
B	6	3	$6 - 3 = 3$	$3^2 = 9$
C	3	4	$3 - 4 = -1$	$(-1)^2 = 1$
D	9	9	$9 - 9 = 0$	$0^2 = 0$
E	1	5	$1 - 5 = -4$	$(-4)^2 = 16$
F	5	7	$5 - 7 = -2$	$(-2)^2 = 4$
G	2	7	$2 - 7 = -5$	$(-5)^2 = 25$
H	7	1	$7 - 1 = 6$	$6^2 = 36$
I	10	8	$10 - 8 = 2$	$2^2 = 4$
J	8	6	$8 - 6 = 2$	$2^2 = 4$
Total	-	-	$\sum d = 1$	$\sum d^2 = 103$

Key Notes on Calculations:
計算要點：

- For Programme A: $d = 4 - 2 = 2$, so $d^2 = 4$.
  節目A：$d = 2$，$d^2 = 4$。

- For Programme G: $d = 2 - 7 = -5$, so $d^2 = 25$.
  節目G：$d = -5$，$d^2 = 25$。

- Sum of d^2: $4 + 9 + 1 + 0 + 16 + 4 + 25 + 36 + 4 + 4 = 103$.
  d^2 總和為：$4 + 9 + 1 + 0 + 16 + 4 + 25 + 36 + 4 + 4 = 103$。

3: Compute Spearman's Rank Correlation (r_s) 計算史皮爾曼等級相關（r_s）

Substitute $\sum d^2 = 103$ and $n = 10$ into the formula:
代入$\sum d^2 = 103$ 和$n = 10$ 到公式：

$$r_s = 1 - \frac{6 \times 103}{10 \times (10^2 - 1)}$$

Step-by-Step Arithmetic:
逐步計算：

1. Calculate denominator inside the fraction:
   計算分數內的分母：
   $n(n^2 - 1) = 10 \times (100 - 1) = 10 \times 99 = 990$

2. Calculate numerator inside the fraction:
   計算分數内的分子：
   $6 \times 103 = 618$

3. Compute the fraction:
   計算分式數值：
   $\frac{618}{990} = 0.6242$ (rounded to 4 decimal places)

4. Final r_s:
   最終r_s值：
   $r_s = 1 - 0.6242 = 0.3758$

4: Adjustment for Tied Ranks (Critical Fix) 並列名次調整（關鍵修正）

The original data has tied ranks (Y gives rank 7 to both F and G). For accuracy, we adjust the formula to account for ties (the standard formula assumes no ties):
原始數據中存在並列名次（Y 對F 和G 都給了等級7）。為準確起見，我們調整公式以考慮並列名次（標準公式假設無並列）：

Tied Rank Adjustment Formula:
並列名次調整公式：

$$r_s = \frac{\sum d^2 - \frac{t_X^3 - t_X}{12} - \frac{t_Y^3 - t_Y}{12}}{\frac{n(n^2-1)}{6} - \frac{t_X^3 - t_X}{12} - \frac{t_Y^3 - t_Y}{12}}$$

Where:
其中：

- t_X = number of tied ranks in X (here, $t_X = 0$—no ties in X),
  $t_X = X$ 中並列名次的數量（此處$t_X = 0$，X 中無並列），

- t_Y = number of tied ranks in Y (here, $t_Y = 2$—two programmes with rank 7).
  $t_Y = Y$ 中並列名次的數量（此處$t_Y = 2$，兩個節目等級7）。

Adjustment Calculation:
調整計算：

1. Tied rank correction for Y:
   Y 的並列名次校正：
   $\frac{t_Y^3 - t_Y}{12} = \frac{2^3-2}{12} = \frac{8-2}{12} = \frac{6}{12} = 0.5$

2. Corrected numerator:
   校正後的分子：
   $\frac{n(n^2-1)}{6} - 0.5 = \frac{990}{6} - 0.5 = 165 - 0.5 = 164.5$

3. Corrected denominator (sum of squared differences with tie adjustment):
   校正後的分母（經並列調整的平方差和）：
   $\sum d^2 - 0.5 = 103 - 0.5 = 102.5$

4. Final adjusted r_s:
   最終調整後的r_s：
   $r_s = 1 - \frac{102.5}{164.5} = 1 - 0.6231 = 0.3769$ (rounded to 4 decimal places)

5: Interpretation 解釋

- Unadjusted $r_s = 0.3758$ (simplified, no tie correction),
  未調整的$r_s = 0.3758$（簡化，無並列校正），

- Adjusted $r_s = 0.3769$ (accurate, accounts for tied ranks).
  調整後的$r_s = 0.3769$（準確，考慮並列名次）。

Both values indicate a moderate positive correlation between the two individuals' rankings (values close to 0 indicate weak correlation; values close to 1 indicate strong positive correlation).
兩個數值均表示兩名人士的排名之間存在中等程度的正相關（接近0 表示弱相關，接近1 表示强正相關）。

Common Errors in Interpreting Correlation Analysis 解讀相關分析時的常見錯誤

As mentioned earlier, correlation analysis is a statistical tool that should be used properly to ensure valid results. Unfortunately, it is sometimes used indiscriminately by management, leading to misleading conclusions. Below are some common errors made in the use of correlation analysis:
如前所述，相關分析是一種統計工具，應正確使用以確保結果有效。不幸的是，管理層有時會不加區分地使用它，導致誤導性的結論。以下是在使用相關分析時常見的一些錯誤：

1. Mistaking Correlation for Causation 將相關誤認為因果

Example 例子

Suppose we study the performance of students in their graduate examinations and their earnings three years after graduation. We might find these two variables are highly and positively correlated. However, we must not overlook that both variables might be influenced by other factors, such as:
假設我們研究學生在畢業考試中的表現與他們畢業三年後收入的關係。我們可能會發現這兩個變數之間存在高度正相關。然而，我們不能忽略這兩個變數可能都受到其他因素的影響，例如：

- Quality of teachers,
  教師的素質，
- Economic and social status of parents,
  父母的經濟與社會地位，
- Effectiveness of the interview process for post-graduation employment,
  畢業後就業面試過程的有效性，
- Other unmeasured contextual factors.
  其他未衡量的背景因素。

If data on these additional factors is available, it is advisable to use *multiple correlation analysis* instead of simple bivariate correlation.
如果有這些額外因素的數據，建議使用多元相關分析而非簡單的雙變量相關。

2. Misinterpreting the Coefficient of Correlation and Coefficient of Determination 誤解相關係數與決定係數

A frequent error is misinterpreting the correlation coefficient (r) and the coefficient of determination (r^2). While this concept was explained earlier, it bears repetition:
一個常見的錯誤是誤解相關係數（r）與決定係數（r^2）的意義。雖然此概念先前已經說明過，但仍值得再次强調：

Example of Misinterpretation 誤解的例子

Suppose a correlation coefficient of $r = 0.7$ is calculated. A common mistake is to claim that this correlation "explains 70% of the total variation in Y." This error becomes clear when calculating the coefficient of determination:
假設計算得到的相關係數為$r = 0.7$。一個常見的錯誤是聲稱此相關「解釋了Y 總變異量的70%」。當計算決定係數時，這個錯誤就會顯而易見：

$$r^2 = (0.7)^2 = 0.49$$

This means only 49% of the total variation in Y is explained by the variation in X—not 70%.
這表示僅有49% 的Y 總變異量可以由X 的變異解釋，而不是70%。

Broader Misinterpretation of r^2 對r^2 的更廣泛誤解

The coefficient of determination is also misused to imply a causal relationship: it is incorrectly interpreted as "the percentage of change in one variable that is due to change in another variable." This is not valid, as r^2 only measures explained variation—not causal impact.
決定係數經常被誤用來暗示因果關係：它被錯誤地解釋為「一個變數的變化中有多少百分比是由另一個變數的變化所造成」。這種解釋並不正確，因為r^2 僅衡量解釋的變異量，而非因果影響。

3. Concluding a Relationship Where None Exists 在無關係時下結論

A third common error is concluding a positive or negative relationship between two variables even when they are *actually unrelated*.
第三個常見的錯誤是，在兩個變數實際上毫無關聯的情況下，仍然得出它們之間存在正相關或負相關的結論。

Example 例子

The age of students and their examination scores may show similar patterns of movement (e.g., both increase slightly over time) but have no true underlying relationship. Their apparent correlation is coincidental, with no common link driving both variables.
學生的年齡與他們的考試成績可能呈現相似的變化模式（例如，兩者都隨著時間略微增加），但實際上並無真正的潛在關係。它們的表面相關是巧合的，沒有共同的聯繫驅動這兩個變數。

Summary 總結

In summary, interpreting the correlation coefficient requires extreme caution. Before concluding a causal relationship, one must consider other relevant factors that might influence the

dependent variable (or both variables). This approach avoids the pitfalls of misinterpreting correlation results. As such, the correlation coefficient is rightly regarded as one of the most widely used—yet also one of the most widely abused—statistical measures.
總結而言，解釋相關係數時必須格外謹慎。在得出因果關係結論之前，必須考慮其他可能影響依變數（或同時影響兩個變數）的相關因素。此做法可避免誤解相關結果的陷阱。因此，相關係數被視為最廣泛使用、但同時也是最常被濫用的統計指標之一。