研究方法設計
- 相關研究法

顧志遠 教授

2023年12月5日

中原大學之建校，本基督愛世之忱，
以信、以望、以愛，致力於中國之高等教育，
旨在追求真知力行，以傳啓文化、服務人類。

相關研究法概論

■ 相關研究法（correlational studies）

□ 相關研究法定義

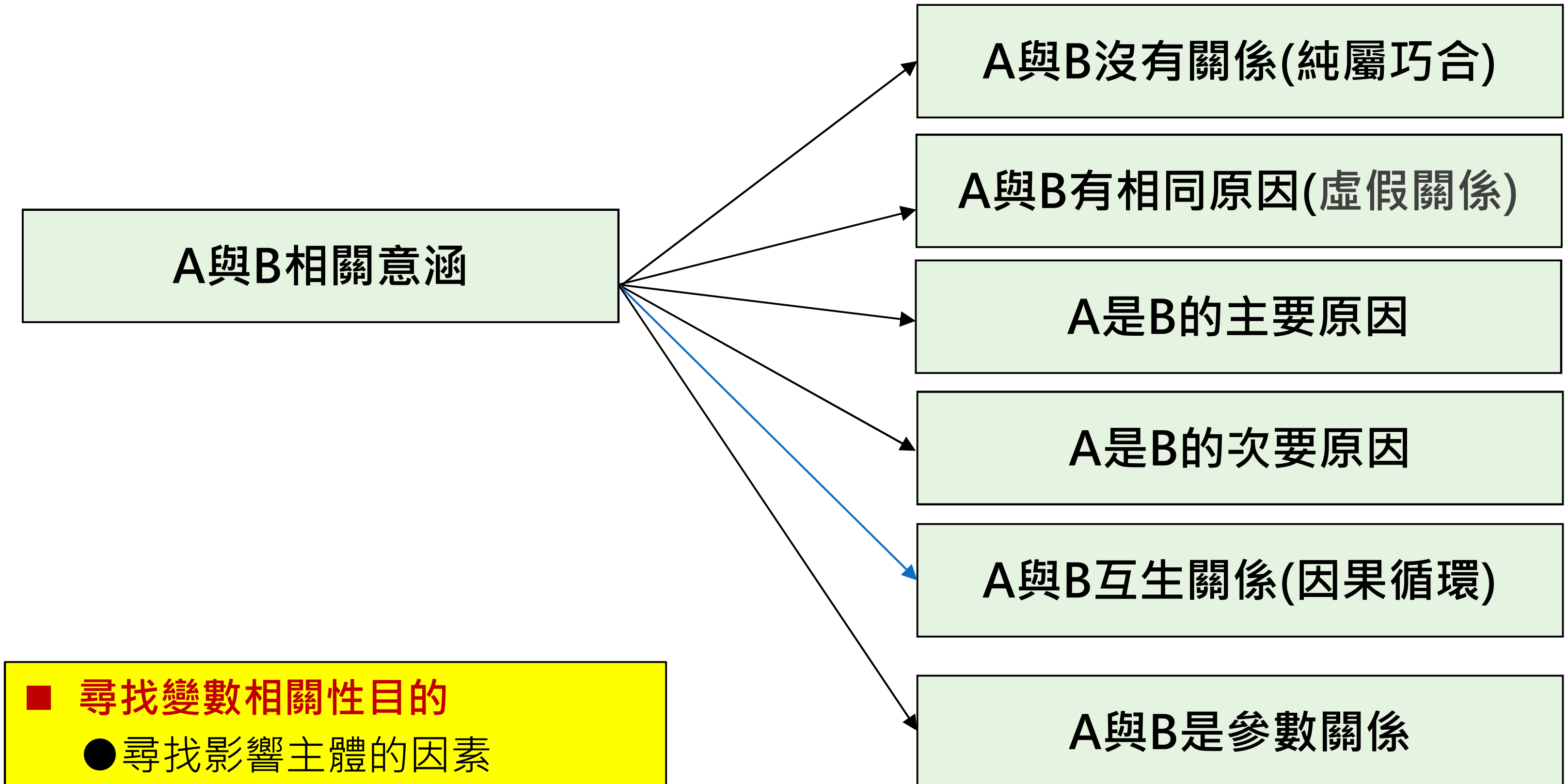
- ✓ 是敘述研究的一種，係採用各種相關統計，探討變項的關聯情形，作為預測或了解因果關係的線索。
- ✓ 凡是經由使用相關係數而探求變項間關係的研究。
- ✓ **相關研究，在設計上並未操縱其中之一個變項，去影響另一變項，同時又沒有控制其他變項，故只能說他們是相關的，不是因果的。**
- ✓ 相關研究用相關係數說明變項間的關係程度，其範圍在1及－1間，負值表負相關，不可視為缺乏相關。

□ 相關研究法的應用

- ✓ 關係研究－適用於探索性研究，在缺乏理論與前人研究下，初步了解一些變數關係，以便做進一步研究的根據。
- ✓ 預測研究－由變項間的相關，從一個變項或某些變項，預測另外的變項。**發生在前之變項，稱預測變項。發生在後之變項，稱效標變項。**但在關係研究中，這兩個變項不需有順序關係。

相關意涵

■ 兩變數有相關之意涵



■ 尋找變數相關性目的

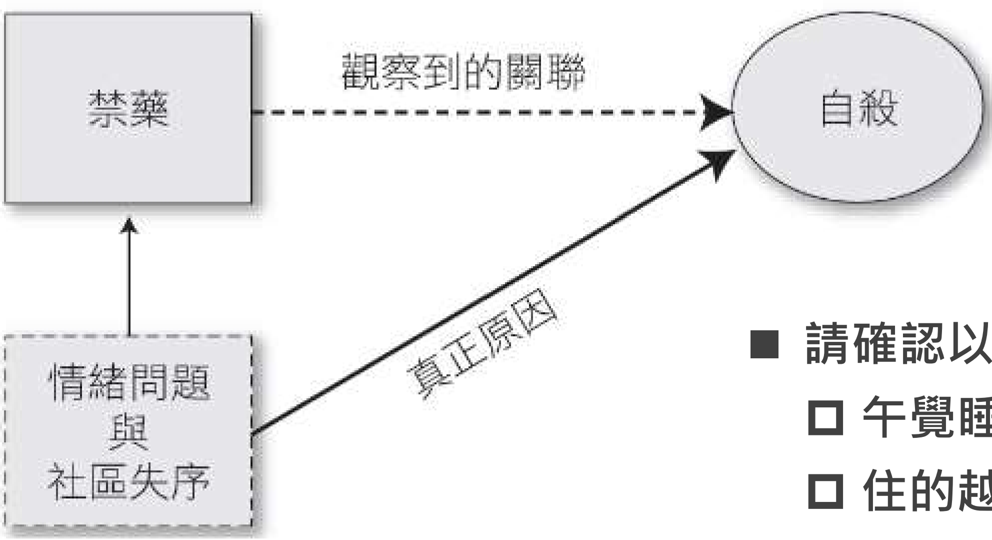
- ●尋找影響主體的因素
- ●建立同一實體的代替指標
- ●建立函數互賴的因次關係
- ●做為因果關係推論的起點

共變關係：冰淇淋銷售量與溺水人數
互生關係：讀書興趣高與讀書成績好
虛假關係：美國人口成長率與澳洲免子成長率
因果關係：智商與學業成績
參數關係：時間與距離

相關意涵

- 虛假性關係是一種錯覺的關係，建立變項關係時要避免虛假關係。源自於未能察覺另有變項是同時造就自變項與依變項的原因。
- 兩個變數的高相關性，是因為第三個變數所達成的，則這個兩個變項稱為虛假關係/偽關係(spuriousness)。

圖 2.10　虛假關係範例——禁藥與自殺間的關係

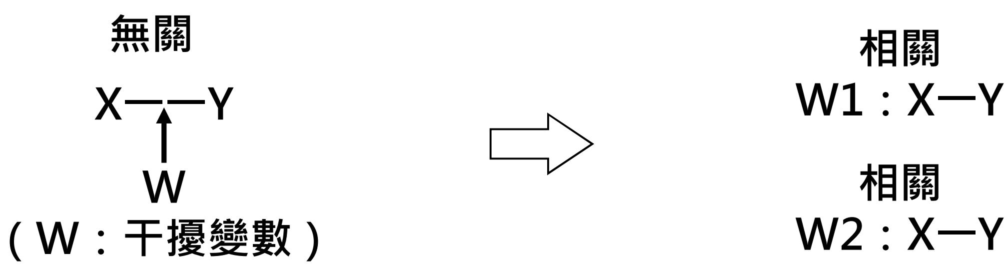


- 請確認以下例子是否為虛假關係？
  - ◻ 午覺睡得越久死亡率越高。
  - ◻ 住的越靠近飛機場死亡率越高。
  - ◻ 冰淇淋銷售量與溺水人數相關。
  - ◻ 城市動物園規模越大犯罪人口越多。

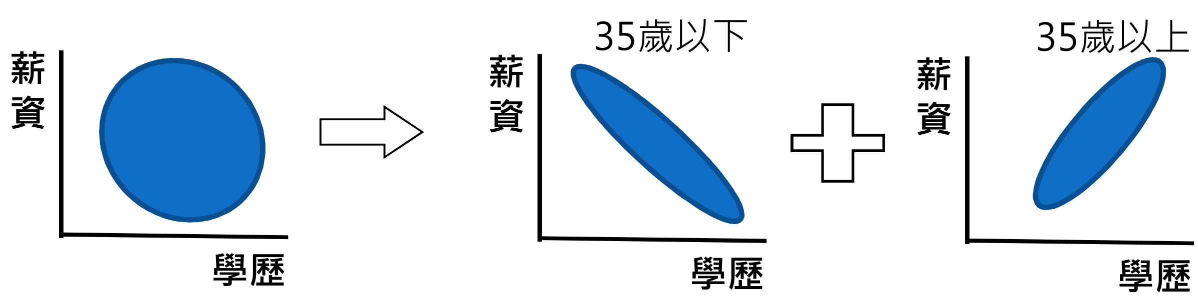
相關意涵

兩變數沒相關之意涵

- 兩個變項真的沒相關。
- 兩個變項是有關的，但看起來沒有關係，因為這兩個變數均與第三個變項有關。則稱此第三個變數為「干擾變數」。而**此兩變項稱為「零偽關係」**。**「干擾變數」亦稱為混淆變項/抑制變項/調節變項/情境變項**。



如學歷與薪水相關與否?實證結果無相關，實際上是受到「年齡」變數干擾。如將資料以某年齡區分為兩群，出現相關性。



辛普森詭論(Simpson's Paradox)

Data Scientist @Trend Micro.

- **在某個條件下的兩組數據，分別討論時都會滿足某種性質，可是一旦合併考慮，卻可能導致相反的結論。**

【**案例一**】有兩家醫院的病患死亡率資料(如下表)，可以看到 A 醫院的的死亡率為15%，B 醫院的死亡率為則為12.5%。如果把剛剛的資料依照「輕重症程度」做細部的分層，可以得到以下的表格，此時可以發現 A 醫院在治療不論重症或是輕症的病人時都有較低的死亡率。在這裡可以發現經過分層後的結果與先前的結果完全相反，這樣子的矛盾現象，稱為「辛普森的悖論」。

	患者人數	死亡人數	死亡率
A 醫院	10000	1500	15%
B 醫院	8000	1000	12.5% 👍

	重症死亡率	輕症死亡率
A 醫院	20% (1400/7000) 👍	3.3% (100/3000) 👍
B 醫院	30% (600/2000)	6.6% (400/6000)

用上述醫院的例子來解釋原因，A 醫院無疑是較好的醫院，但由於 A 醫院是國內頂尖的醫院，因此病情較為嚴重的患者會優先至 A 醫院就診。而重症患者的死亡率天生就遠高於輕症患者，進而拉高了 A 醫院的整體死亡率。於是乎當不考慮病患的病情而只看整體的死亡率時，會誤以為 B 醫院的醫療醫療水平較高。

辛普森詭論(Simpson's Paradox)

- 在某個條件下的兩組數據，分別討論時都會滿足某種性質，可是一旦合併考慮，卻可能導致相反的結論。

【案例二】

電機工程		
	男	女
通過	30	10
拒絕	30	10
總和	60	20

男女申請通過比例皆50%

＋

企業管理		
	男	女
通過	5	10
拒絕	15	30
總和	20	40

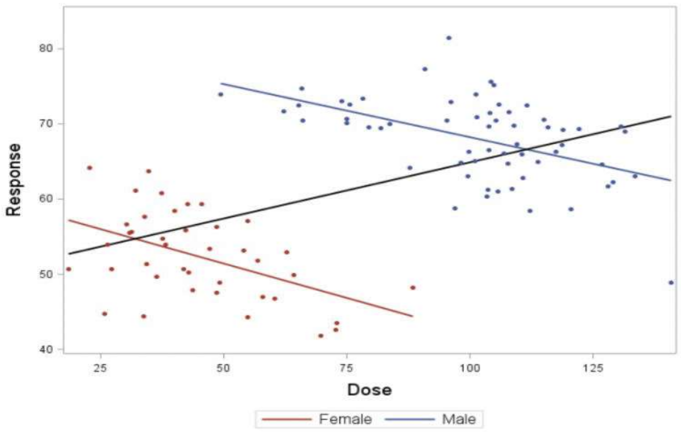
男女申請通過比例皆33%

＝

合併			
	男	女	總和
通過	35	20	55
拒絕	45	40	85
總和	80	60	140

女申請通過比例=33%
男申請通過比例=44%

【案例三】



不同性別分開看時，Dose & Response皆呈現負相關!!

合併一起看時，觀察Dose & Response卻呈現正相關!!

➔**性別同時會影響Dose及Response!!**

當潛在(干擾)變數存在時，觀察到的關聯有可能是誤導的，辛普森詭論只是這項事實的一種極端形式!!

相關意涵

- 兩變數是非線性相關，若用線性相關公式計算， 呈現不相關。

  無關 X — Y ⇒ 相關 Y —F(X) ⇒ X,Y 非線性相關係

- 兩變數因次關係，若用線性相關公式計算， 呈現不相關。

  無關 X — Y ⇒ 相關 X —Y^n ⇒ X,Y 因次關係

- 兩個變項間相關，但應了解其是否為直接相關或有中介變數。

  無關 X — Y ⇒ 相關 X—W 相關 W—Y ⇒ W：中介變數

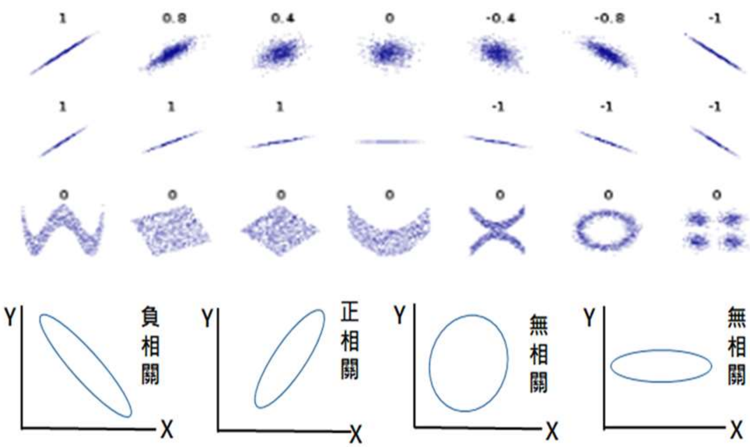
兩變數間關係的涉及因素

- 正相關(＋)或負相關(－)
- 相關性的強度 %
- 對稱或不對稱 X→Y，Y→X， X$\boxed{\leftrightarrow}$Y
- 企業研究最有興趣的非對稱關有：刺激→反應，屬性→傾向，傾向→行為，屬性→行為。
- 線性或非線性

相關意涵

■ 相關係數的範圍

- ◘ 研究用相關係數說明變項間的關係程度，其範圍在1及 - 1間，負值表負相關，不可視為缺乏相關。
- ◘ 正相關：一個變項增加(減少)，另一個變項隨之增加(減少)。
- ◘ 負相關：一個變項增加(減少)，另一個變項隨之減少(增加)。
- ◘ 零相關：一個變項增加(減少)，另一個變項沒有隨之變化。



■ 相關係數的解釋

- ◘ 相關係數的意義應視樣本大小而定，如要達到不是零相關99%的顯著水準，樣本數為5，則其相關係數要大於0.96，若樣本數為52人，則其相關係數只須大於0.35，若樣本數為102人，則相關係數只須0.25。
- ◘ 相關係數不等於百分比，γ2表示某變項可由另一變項決定的比率。
- ◘ 相關係數不是等距或等比變項，即不可解釋0.5是0.25的兩倍及0.9和0.8的相關差與0.2和0.3的相關差相同。通常要達到（正）更高或（負）更低的相關度，愈為困難。
- ◘ 相關不等於因果關係，可進一步實驗的方式，確定因果關係。
- ◘ 相關係數大可和資料的變異程度有關。

$$E(R^2) \approx \frac{\hat{B}_1^2 S_{XX}}{\hat{B}_1 S_{XX} + \sigma^2}$$

$$S_{XX}\uparrow \rightarrow E(R^2)\uparrow$$

$$S_{XX}\downarrow \rightarrow E(R^2)\downarrow$$

相關公式種類 - 簡單相關

■ SPSS-I-008-01-相關分析法概覽

相關分析法	變項數	方法說明
簡單相關	**1—1**	解釋**一個**自變項和**一個**依變項之間關係的方法。
偏相關 (淨相關)	**1 (-1)—1 (-1)**	在**控制**第三個變項(z變項)的情況下，用來分析兩個變項(x變項和y變項)之間關係的方法。也就是兩個變項(x變項和y變項)**同時去除**z變項的影響後，所得的相關程度稱為**淨相關**。
		【研究實例】 1. 員工工作滿意、組織承諾與離職意圖關係的統合分析修正模式 2. 概念構圖學習成效相關影響因素之探討
部份相關	**1—1 (-1)**	分析兩個變項(x變項和y變項)之間的**單純相關**。如果在計算排除效果時，**只處理**第三變項(z變項)與x、y變項當中任一個的相關時求得的相關係數稱為**部份相關**，或稱為**半淨相關**。
複相關 (多元相關)	**多—1**	綜合**一組**自變項(2個以上的變項，x變項)對**一個**依變項(y變項)的整體影響力。
		【研究實例】 1. 台灣地區宅配人員職業焦崩對工作滿足、組織承諾與離職傾向關係之研究
典型相關	**多—多**	用來解釋**一組**自變項 (2個以上的變項，x變項)與另**一組**依變項(2個以上的變項，y變項)之間的關係的分析方法。
		【研究實例】 1. 產業別財務變數差異研究

相關公式種類 - 簡單相關

分析方法	符號	變 項1	變 項 2	備註	目　的
積差相關	γ	連續變數	連續變項	用於兩個變項均使用等距或等比分數。	分析兩個變項間的直線關係
等級相關	ρ	等級變項	等級變項	又稱Spearman相關係數，適用兩個變數均為等級或順序分數。	
肯氏係數	τ	等級變項	等級變項	適用在樣本人數10以下之兩變數為等級分數。	
肯氏和諧係數	ω	等級變項	等級變項	適用於分析多位評分者評分一致性，亦適用於等級評分。	分析評分者的一致性
二系列相關	Γbis	人為二分類變項	連續變項	適用在變項為常態的連續分數，變項為二分類別評分，常用於分析試題鑑別力	分析試題的鑑別力
點二系列相關	Γp.bis	真正二分類變項	連續變項	適用於一個變項是連續分數，另一變項為真正的二分類評分。	
四分相關	γt	人為二分類變項	人為二分類變項	適用在兩個變數均為常態分配的連續分數，並且都用人為方法分成兩個類別。	兩個變數均可以二分時使用
Φ相關	Φ	真正二分類變項	真正二分類變項	適用於兩個變數均為真正的二分變項。適用此法的例子不多，故使用不多。	分析試題間的相關
列聯相關	C	二個以上類別	二個以上類別	適用於兩個變項的分類，超過兩個以上的類別。	兩個變數均分成若干類時使用
相關比	η	連續變數	連續變數	非直線相關，即兩變項在此種程度前，兩者是同方向增減，但之前呈反向關係	分析非直線的相關

無母數分析方法

- **無母數分析方法(檢定)：**
  無法符合有母數分析所設計的方法；常使用符號(正負)或排序(大小順序)取代測量數值，或使用各分類的次數以進行統計分析。適用於類別、序位尺度資料分析與資料分布未知的情況。

- **無母數分析方法優點：**
  - ✓ 母群體分布未知或不是常態分布，或是樣本數不夠大時皆可使用。是無母數分析方法的最大優點。
  - ✓ 計算簡單且快速。
  - ✓ 雖然在母群實際上為常態分配時，較有母數分析方法不易得到顯著結果；但在母群體不是常態分布時，無母數分析之檢力較有母數分析高。

- **無母數分析方法缺點：**
  - ✓ 只使用資料的符號、排序等特性，浪費了數值之集中趨勢、分散性及分佈所提供的資訊。
  - ✓ 針對常態分布資料如果仍進行無母數分析，將使檢力降低。

- **當欲檢定的資料不符合有母數分析法之假設前提時才建議使用無母數分析法，為一種互補的統計方法，而非用於取代有母數分析法。**

相關公式種類 - 簡單相關

■ SPSS-I-008-01-相關分析法概覽

		連續變項	次序變項		類別變項			
			真正次序	人為次序	類別變項	真正二分	人為二分	真正二分 (多項)
連續變項		Pearson積差 相關/淨相關/ 相關比	點二系列關	二系列相關	點二系列相關	點二系列相關	二系列相關	
次序變項	真正次序	點二系列相關	Spearman 等級相關 Kendall 和諧係數					
	人為次序	二系列相關		Kendall 等級係數				
類別變項	類別變項	點二系列相關			Kappa 一致性係數			
	真正二分	點二系列相關				ψ相關		
	人為二分	二系列相關					四分相關	
	真正二分 (多項)							列聯相關

Pearson積差相關(γ)	
變項一：連續分數(等距、比率變項)	
變項二：連續分數(等距、比率變項)	
相關公式	$r_{xy} = \frac{\sum Z_x Z_y}{N} = \frac{C_{xy}}{S_x S_y} = \frac{\sum (x-\bar{x})(y-\bar{y})}{N S_x S_y}$ $= \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{N}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{N}}}$
適用情況	◘ 等距或等比量尺測量到的連續分數
使用範例	◘ 工作時數與收入的關係。 ◘ 國語成績與數學成績的關係。

Spearman等級相關(ρ)	
變項一：次序變項	
變項二：次序變項	
相關公式	(1)二個變項沒有一致的等級： $W = \frac{S}{\frac{1}{12} \cdot K^2 \cdot (N^3 - N)}$ $S = \sum R_i^2 - \frac{(\sum R_i)^2}{N} = \sum (R_i - R)^2$ R：評分等級； K：評分者人數； N：被評分的人數或作品件數。 (2)二個變項有一致的等級： $W = \frac{S}{\frac{1}{12} \cdot K^2 \cdot (N^3 - N) - K \sum T}$ $S = \sum R_i^2 - \frac{(\sum R_i)^2}{N} = \sum (R_i - R)^2$ $\sum T = \frac{t^3 - t}{12}$ ；$(K > 3)$ R：評分等級；K：評分者人數； N：被評分的人數或作品件數；t：表示得到相同等第的人數。
適用情況	◘ 適用於評分者間信度(interjudge reliability)。 ◘ 檢定多位評分者對N件作品評分一致性。
使用範例	◘ 五位評審評10位學生的美術作品了解其評分是否一致。

Kendall和諧係數(ω)	
變項一：次序變項	
變項二：次序變項	
相關公式	(1)二個變項沒有一致的等級： $r_s = 1 - \frac{6\sum D^2}{N(N^2-1)}$ ，D：為二變數對稱之等級差 (2)二個變項有一致的等級： $r_s = \frac{\sum x^2 + \sum y^2 - \sum D^2}{2\sqrt{\sum x^2 \sum y^2}}$ 其中， $\sum x^2 = \frac{N^3 - N}{12} - \sum Tx$ $\sum y^2 = \frac{N^3 - N}{12} - \sum Ty$ $\sum T = \frac{t^3 - t}{12}$ ，t：表示得到相同等第的人數。
適用情況	◻ 二個評分者評N件作品，以了解評分是否一致。 ◻ 一個評分者，先後二次評N件作品，以了解評分是否一致。
使用範例	◻ 二位評審評10位學生的美術作品，了解其評分是否一致。

Kendall等級係數(τ)	
變項一：人為次序變項	
變項二：人為次序變項	
相關公式	$\tau = \frac{S}{\frac{1}{2}N(N-1)}$ S：等第次序量數 N：被評者的人數或作品件數
適用情況	◻ 簡便好用。 ◻ 樣本人數在10人以下，適用此法。 ◻ 比Spearman等級相關更有優點。
使用範例	◻ 二位評審評10位學生的美術作品，了解其評分是否一致。

二系列相關(Γbis)	
變項一：人為二分變項	
變項二：連續變項	
相關公式	$r_{bis} = \dfrac{\overline{X}_p - \overline{X}_q}{S_t} \cdot \dfrac{p \cdot q}{y}$ $\overline{X}_p$: 表第一類之平均數； $\overline{X}_q$: 表第二類之平均數； S_t: 表全體分數之標準差； p：表第一類人數之百分比；q：表第二類人數之百分比。
適用情況	☐ 適用兩變項均為常態的連續分數，但其中一變項被分為兩個類別時，此法所得係數比積差相關為高，若變項不是常態分配時，則rbis係數可能超過1以上。 ☐ 收集的資料都是連續變項時，最好採用積差相關法分析。 ☐ 測驗編製時，此法往往被用來做試題分析，以決定鑑別力。
使用範例	☐ 分析代數成績及格與否和數學性向的關係。其中代數成績屬常態分配的連續變數，但人為分成及格與否兩類。

點二系列相關(Γp.bis)	
變項一：真正二分變項(名義變項)	
變項二：連續變項	
相關公式	$r_{pq} = \dfrac{\overline{X}_p - \overline{X}_q}{S_t} \cdot \sqrt{pq}$ $\overline{X}_p$: 表第一類之平均數； $\overline{X}_q$: 表第二類之平均數； S_t: 表全體分數之標準差； p：表第一類人數之百分比；q：表第二類人數之百分比。
適用情況	☐ 適用於一變項為連續變項，另一變項為真正的二分類變項； ☐ 此法分析比二系列相關為低，且不超過1。 ☐ 可用來做試題分析，以決定鑑別力
使用範例	☐ 性別與智力的關係。 ☐ 性別與收入的關係

四分相關(γt)	
變項一：人為二分名義變項	
變項二：人為二分名義變項	
相關公式	$r_{tet} = \cos\left(\dfrac{180^\circ}{1+\sqrt{\dfrac{BC}{AD}}}\right)$
適用情況	□ 適用於兩變項為常態分配的連續分數但兩變項皆是人為分二類。 □ 比積差相關不穩定。 □ 標準誤比積差相關大。
使用範例	□ 工作動機(高動機、低動機)與工作績效(達到標準、未達標準，或是成功失敗)的關係。 □ 學生的學業成績(及格、不及格)和智商(高、低)的關係。

ψ相關(Φ)	
變項一：真正二分變項	
變項二：真正二分變項	
相關公式	$\phi = \dfrac{p_{xy} - p_x p_y}{\sqrt{p_x q_x}\sqrt{p_y q_y}} = \dfrac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$
適用情況	□ 適用於二個真正二分變項。 □ 比四分相關為低，但較其穩定，其顯著水準易於考驗。
使用範例	□ 性別與使用左右手的習慣。 □ 性別與讀大學的關係

Kappa一致性係數(K)	
變項一：類別變項(人為分類)	
變項二：類別變項(人為分類)	
相關公式	Kappa 一致性係數的概念是，討論評分者的評分一致性的百分比與評分者理論上評定的次數百分比之比率。 $K=\frac{P(A)-P(E)}{1-P(E)}$ P(A)為 K 位評分者評分一致的百分比， $P(A)=\left[\frac{1}{NK(K-1)}\sum_{i=1}^{N}\sum_{j=1}^{m}n_{ij}^{2}\right]-\frac{1}{K-1}$ N：總人數；K：評分者人數；m：評定類別；n：細格資料。
適用情況	◻ Kendall和諧係數用於評分者可對評分對象定出等第； ◻ Kappa一致性係數一用於評分者無法對評分對象定出等第，僅能將其分類。
使用範例	◻ 二位資優鑑定委員將一群資優學生依據診鑑工具的測量結果，將其歸屬在不同類別的資優班級中。

列聯相關(C)	
變項一：二分或二個以上分類的名義變項	
變項二：二分或二個以上分類的名義變項	
相關公式	$C=\sqrt{\frac{\chi^2}{N+\chi^2}}$ C：最大值為$\sqrt{\frac{m-1}{m}}$; N為總人數。
適用情況	◻ 適用於二個分類變項的類別有二個或超過兩個以上時。 ◻ 列聯相關與卡方考驗有密切關係，它可以由卡方考驗的資料直接求得
使用範例	◻ 不同年級的學生(四、五、六年級)對於實施課後社團活動的意見(同意、無意見、不同意)。 ◻ l不同教育程度(大學、高中、國中、國小)對體罰的看法(贊成、不贊成、無意見) ◻ l網友的上網時段(白天、晚上、假日)和上網動機(購物、交友、找資料)的關係。

曲線相關或相關比(η)	
變項一：連續變項	
變項二：連續變項	
相關公式	$\eta_{xy} = \sqrt{\frac{SS_b}{SS_t}}$
適用情況	◻ 非直線相關，其特性在於開始時，X變項增加或減少，Y變項也增加或減少；但到了某程度後，兩者的增減方向正好相反，這種關係即可稱為曲線相關或相關比。。 ◻ 如果兩變項屬於非直線相關，採用直線相關分析將導致低估變項的關係程度。
使用範例	◻ 補習時間和考試成績。 ◻ 工作效率和焦慮程度。

淨相關($r_{12.3}$)	
變項一：連續變項	
變項二：連續變項	
相關公式	$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$ 其顯著性考驗 $t = \frac{r_{12.3}}{\sqrt{\frac{1 - r_{12.3}^2}{N-3}}}$
適用情況	◻ 去除和二個連續變項都有關連的影響因素後，再求得二個變項之間純粹的關係。
使用範例	◻ 去除智力的影響，再求數學和國文成績的關係。 ◻ 去除動機的影響，再求工作表現和收入的關係。

相關公式 - 淨相關(Partial Correlation)

- 淨相關是指在控制第三個變項(z變項)的情況下，用來分析兩個變項(x變項和y變項)之間關係的方法。也就是兩個變項(x變項和y變項)同時去除z變項的影響後，所得的相關程度稱為淨相關。
- 當研究者想知道二個連續變項在第三個變項出現時，會對原先的二個變項產生什麼樣的變化，即可用淨相關分析。

【研究問題】

研究者依序控制**信念**、**素養**和**時間**，進一步想了解**領導**、**行政**、**環境**三個變項間彼此間的關係，則**領導**、**行政**、**環境**三者間的一、二、三階淨相關為何？

SPSS-I-008-02-淨相關-資料範例.sav [資料集2] - SPSS Statistics Data Editor

檔案(F) 編輯(E) 檢視(V) 資料(D) 轉換(T) 分析(A) 統計圖(G) 公用程式(U) 增益集(O) 視窗(W) 說明(H)

23 : 環境總分 22.0

	編號	性別	年齡	學歷	職務	時數	班數	知覺總分	預勢總分	導引總分	機械總分	統整總分	適應總分	創新總分	層級總分	信念總分	素養總分	時間總分	領導總分	行政總分	環境總分	影響總分
1	001	2	3	1	4	3	4	26	27	23	29	22	23	24	174	27	30	23	17	22	23	142
2	002	1	3	1	3	4	4	17	20	19	21	19	22	17	135	19	20	18	18	21	20	116
3	003	1	2	1	3	3	4	10	12	12	16	11	12	12	85	12	13	13	11	14	14	77
4	004	2	2	1	3	2	4	22	27	23	22	21	24	20	159	22	24	18	19	23	25	131
5	005	2	3	2	1	4	4	18	22	19	17	14	16	17	123	21	22	16	15	23	23	120
6	006	2	3	1	3	3	4	22	24	22	21	20	22	20	151	21	22	20	17	21	22	123
7	007	2	3	3	1	4	4	16	21	20	19	16	17	15	124	19	23	22	19	22	21	126
8	008	2	3	1	1	4	4	18	22	21	17	16	16	13	123	16	18	26	13	15	13	101
9	009	2	3	3	4	1	4	21	22	20	21	18	19	20	141	18	19	18	13	15	15	98

控制變項個數	淨相關階層	淨相關符號
0	零階淨相關(簡單相關)	r_{xy}
1	一階淨相關	$r_{xy\cdot 1} = \dfrac{r_{xy} - r_{y1} \cdot r_{x1}}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{x1}^2}}$ $r_{xy\cdot 1}$：x、y變項在同時排除控制變項1的影響後，x、y變項之間的淨相關。 r_{xy}：x、y變項的積差相關(零階淨相關)。 r_{y1}：y變項和控制變項1的積差相關(零階淨相關)。 r_{x1}：x變項和控制變項1的積差相關(零階淨相關)。
2	二階淨相關	$r_{xy\cdot 12} = \dfrac{r_{xy\cdot 2} - r_{x1\cdot 2} \cdot r_{y1\cdot 2}}{\sqrt{1 - r_{x1\cdot 2}^2}\sqrt{1 - r_{y1\cdot 2}^2}}$

相關公式 - 淨相關(Partial Correlation)

【一階淨相關】排除控制變項「信念」的影響

- 「領導」、「行政」淨相關為 0.667 (p=.000)；
- 「領導」、「環境」淨相關為 0.531 (p=.000)；
- 「行政」、「環境」淨相關為 0.798 (p=.000)；

控制變數			領導總分	行政總分	環境總分
信念總分	領導總分	相關	1.000	.667	.531
		顯著性 (雙尾)	.	.000	.000
		df	0	157	157
	行政總分	相關	.667	1.000	.798
		顯著性 (雙尾)	.000	.	.000
		df	157	0	157
	環境總分	相關	.531	.798	1.000
		顯著性 (雙尾)	.000	.000	.
		df	157	157	0

【二階淨相關】排除控制變項「信念」、「素養」的影響

- 「領導」、「行政」淨相關為 0.675 (p=.000)；
- 「領導」、「環境」淨相關為 0.537 (p=.000)；
- 「行政」、「環境」淨相關為 0.798 (p=.000)；

控制變數			領導總分	行政總分	環境總分
信念總分 & 素養總分	領導總分	相關	1.000	.675	.537
		顯著性 (雙尾)	.	.000	.000
		df	0	156	156
	行政總分	相關	.675	1.000	.798
		顯著性 (雙尾)	.000	.	.000
		df	156	0	156
	環境總分	相關	.537	.798	1.000
		顯著性 (雙尾)	.000	.000	.
		df	156	156	0

【三階淨相關】排除控制變項「信念」、「素養」、「時間」的影響

- 「領導」、「行政」淨相關為 0.700 (p=.000)；
- 「領導」、「環境」淨相關為 0.565 (p=.000)；
- 「行政」、「環境」淨相關為 0.799 (p=.000)；

控制變數			領導總分	行政總分	環境總分
信念總分 & 素養總分 & 時間總分	領導總分	相關	1.000	.700	.565
		顯著性 (雙尾)	.	.000	.000
		df	0	155	155
	行政總分	相關	.700	1.000	.799
		顯著性 (雙尾)	.000	.	.000
		df	155	0	155
	環境總分	相關	.565	.799	1.000
		顯著性 (雙尾)	.000	.000	.
		df	155	155	0

相關公式 - 淨相關(Partial Correlation)

- **狀況1： $r_{yx\cdot z}$ 與 r_{yx} 差不多**

  表示控制變項對此兩變項之關係並無影響，亦即此兩變項間有直接關係。

- **狀況2: $r_{yx\cdot z} < r_{yx}$ (小很多時)**

  - X、Y間是虛假關係(Spurious relationship): 即Z為X與Y的共同因，X與Y間實際上無關係。

    Z

    ↙ ↘

    X — Y

  - Z為X與Y間的中介變項

    Z

    ↗ ↘

    X — Y

  當 $r_{yx\cdot z}$小於 r_{yx}極多，但 $r_{yx} \neq 0$時，必須將Z放入因果關係中。但是淨相關無法說明Z是共同因或是為中介變項，必須要從時間先後或是從理論推斷。

- **狀況3: $r_{yx\cdot z} > r_{yx}$ (大很多時)**

  表示控制項獨立影響Y (即XY間與ZY間的關係為相互獨立) 。但加入Z後，因Y與Z共同變異部分已排除（或是說因為 Y 部份的變異量已被 Z 所解釋），X只需解釋Y變異量中尚未被Z解釋的部份，故 r_{yx}升高。

  X ↘

  Y

  Z ↗

相關公式 - 部分相關(Part Correlation)

- 分析兩個變項(x變項和y變項)之間的單純相關。如果在計算排除效果時，只處理第三變項(z變項)與x、y變項當中任一個的相關時，求得的相關係數稱為部份相關(part correlation)，或半淨相關(semipartial correlation)。
- 部份相關係數由於部份相關只處理兩個變項(x變項和y變項)其中的一個變項，其係數符號的表達方式會依處理對象有不同的表示方法：
  (1) $r_{x(y \cdot z)}$：表示x、y變項的部份相關係數，且排除變項**y和z的相關**；
  (2) $r_{y(x \cdot z)}$：表示y、x變項的部份相關係數，且排除變項**x和z的相關**；

【研究問題】

排除「信念」和「素養」的相關，「信念」和「層級」的部分相關為何？

(1)「信念」和「層級」淨相關為 .732 (p=.000)，達顯著，顯示二者仍有顯著高相關，但相關程度比簡單相關(零階) .881小；

(2)「素養」和「層級」簡單相關(零階)為 .721 (p=.286)，未達顯著；

(3)「信念」和「層級」部份相關為 .507 (p=.000)，達顯著；反觀，「素養」和「層級」的部份相關為 .040 (p=.286)，此值與零階的簡單相關係數 .721小很多；顯示「信念」和「層級」的相關，在排除「信念」和「素養」的相關後，所呈現的部份相關效果很明顯。

$$r_{x(y \cdot z)} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{1 - r_{yz}}}$$

係數[a]

模式		未標準化係數		標準化係數			相關		
		B 之估計值	標準誤差	Beta 分配	t	顯著性	零階	偏	部分
1	(常數)	20.137	5.762		3.495	.001			
	信念總分	5.608	.416	.829	13.479	.000	.881	.732	.507
	素養總分	.405	.379	.066	1.070	.286	.721	.085	.040

a. 依變數: 層級總分

相關公式 - 複相關(Multiple Correlation Analysis)

- **複相關(Multiple Correlation Analysis)，又稱多元相關。複相關是綜合一組自變項(2個以上的變項，x變項)對一個依變項(y變項)的整體影響力。**
- **複相關係數**

$R_{y\cdot12}$：表示依變項y對兩個自變項1、2的複相關係數。

$R_{y\cdot12}^2$：表示自變項1、2對依變項y的變總變異之解釋量。

【**研究問題**】

影響教師數位教學技能的因素「信念」、「素養」、「時間」、「領導」、「行政」、「環境」，對教師數位教學技能「層級」的複相關及解釋量為何？

$$R_{y\cdot12}=\sqrt{\frac{r_{y1}(r_{y1}-r_{y2}r_{12})+r_{y2}(r_{y2}-r_{y1}r_{12})}{1-r_{12}^{\ 2}}}$$

(1) 複相關係數R值 = 0.885 (p = .000)，達顯著，表示「信念」、「素養」、「時間」、「領導」、「行政」、「環境」對教師數位教學技能「層級」有高度的相關；

(2) 調過後的R平方 = 0.774，表示「信念」、「素養」、「時間」、「領導」、「行政」、「環境」對教師數位教學技能「層級」的總變異解釋量為77.4%。

模式摘要

模式	R	R平方	調過後的R平方	估計的標準誤	變更統計量				
					R平方改變量	F改變	df1	df2	顯著性F改變
1	.885[a]	.782	.774	11.313	.782	91.732	6	153	.000

a. 預測變數:(常數), 環境總分, 時間總分, 素養總分, 領導總分, 信念總分, 行政總分

兩個變數相關公式 — 評價診斷試驗

■ 生物統計學-常見評價診斷試驗真實性的指標包括靈敏度(真陽性率)、特異度(真陰性率)、假陽性率、假陰性率、似然比。

一、真陽性、假陽性、真陰性、假陰性

指篩檢或診斷試驗中所得到的數值與實際情況符合的程度，稱作真實性(validity)。

	生病(Disease)	健康(non Disease)	
陽性	a (真陽性)	b (假陽性)	a+b
陰性	c (假陰性)	d (真陰性)	c+d
	a+c	b+d	a+b+c+d

三、Likelihood Ratios (似然比)：

病患中出現某種試驗結果的概率與非病患中出現該結果的概率之比，說明病患出現該結果的機率是非病患的多少倍。似然比是一個綜合性評價指標，不受患病率的影響，在選擇診斷試驗時應選擇陽性似然比高的方法。

陽性似然比	意義
>10	Strong evidence to rule in disease
5–10	Moderate evidence to rule in disease
2–5	Weak evidence to rule in disease
0.5–2.0	No significant change in the likelihood
0.2–0.5	Weak evidence to rule out disease
0.1–0.2	Moderate evidence to rule out disease
<0.1	Strong evidence to rule out disease

	定義	公式	意義
陽性似然比 (positive likelihood ratio，+LR)	病患中某試驗出現陽性結果的機率是非病患的多少倍---即真陽性率與假陽性率之比	[a/(a+c)]/[b/(b+d)]	比值愈大，患病的機率愈大；比值愈小，基本可以排除疾病
陰性似然比 (negative likelihood ratio，-LR)	病患中某種試驗出現陰性結果的機率是非病患的多少倍—即假陰性率與真陰性率之比	[c/(a+c)]/[d/(b+d)]	比值越小，試驗診斷的價值愈高

兩個變數相關公式 — 評價診斷試驗

二、靈敏度與特異度

上述表格是用來計算靈敏度與特異度，二者是統計學中用來表示二項分類測試特徵的數據。靈敏度只與病例組有關，理想的試驗靈敏度應為100%；特異度只與非病例組有關，理想的試驗特異度應為100%。

	定義	公式
靈敏度 (真陽性率) (Sensitivity)	將有病者診斷為陽性結果的比率= 真陽性/生病	a / a+c
特異度 (真陰性率) (Specificity)	將健康者診斷為陰性結果的比率= 真陰性/健康	d / b+d
假陽性率 (false positive rate，Fpr) (誤診率) (第一類錯誤)	實際無疾病，但根據診斷試驗卻被定為有病的機率	b / b+d = 1 - 特異度
假陰性率 (false negative rate，Fnr) (漏診率) (第二類錯誤)	實際有疾病，但根據診斷試驗卻被定為非病者的機率	c / a+c = 1 - 靈敏度

	意義
靈敏度 (真陽性率) (Sensitivity)	當靈敏度高的診斷試驗結果為陰性，是未罹患此疾病一項相當可靠指標
特異度 (真陰性率) (Specificity)	在特異性高的診斷試驗結果為陽性，即表示有病
假陽性率 (false positive rate，Fpr) (誤診率) (第一類錯誤)	特異度越高，誤診愈少，理想的診斷試驗假陽性率=0
假陰性率 (false negative rate，Fnr) (漏診率) (第二類錯誤)	靈敏度越高，漏診愈少，理想的診斷試驗假陰性率=0

四、Youdens index(r)：

r=(靈敏度+特異度) - 1 = 1 - (假陽性率+假陰性率)，用於兩個診斷方法的比較，理想的正確診斷指數為100%。

兩個變數相關公式 — dissimilarity measures

Binary Data

There are a surprisingly large number of dissimilarity measures available for binary data, for example SPSS v8.0 lists 27 different ones. All of these binary-based measures use two or more of the values obtained from a simple 2 x 2 matrix of agreement. Note that n = a + b + c + d.

	+	-
+	a	b
-	c	d

Euclidean distance.	SQRT(b+c)	The square root of the sum of discordant cases, minimum value is 0, and it has no upper limit.
Squared Euclidean distance	b+c.	The sum of discordant cases, minimum value is 0, and it has no upper limit.
Pattern difference	bc/(n**2)	Dissimilarity measure for binary data that ranges from 0 to 1.
Variance.	(b+c)/4n	Ranges from 0 to 1
Simple matching	(a+d)/n	This is the ratio of matches to the total number of values. Equal weight is given to matches and nonmatches.
Dice.	2a/(2a + b + c)	This is an index in which joint absences are excluded from consideration, and matches are weighted double. Also known as the Czekanowski or Sorensen measure.
Lance and Williams	(b+c)/(2a+b+c)	Range of 0 to 1. (Also known as the Bray-Curtis nonmetric coefficient.)
Nei & Lei's genetic distance	2a/[(a+b)+(a+c)]	
Yule coefficient	(ad-bc)/(ad+bc)	This index is a function of the cross-ratio for a 2 x 2 table and is independent of the marginal totals. It has a range of -1 to 1.

兩個變數相關公式 — 混淆矩陣(confusion matrix) 維基百科

- 在機器學習領域和統計分類問題中，混淆矩陣具有兩個維度的(實際和預測)列聯表，通過這個矩陣可以方便地看出機器是否將兩個不同的類混淆。

Sources: [20][21][22][23][24][25][26][27]

		Predicted condition			
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP), hit**	**False negative (FN), type II error, miss, underestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
Actual condition	**Negative (N)**	**False positive (FP), type I error, false alarm, overestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2 PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

兩個變數相關公式 — 混淆矩陣(confusion matrix)

$$S_{\text{JACCARD}}=\frac{a}{a+b+c} \quad (1)$$

$$S_{\text{DICE}}=\frac{2a}{2a+b+c} \quad (2)$$

$$S_{\text{CZEKANOWSKI}}=\frac{2a}{2a+b+c} \quad (3)$$

$$S_{\text{3W-JACCARD}}=\frac{3a}{3a+b+c} \quad (4)$$

$$S_{\text{NEI\&LI}}=\frac{2a}{(a+b)+(a+c)} \quad (5)$$

$$S_{\text{SOKAL\&SNEATH-I}}=\frac{a}{a+2b+2c} \quad (6)$$

$$S_{\text{SOKAL\&MICHENER}}=\frac{a+d}{a+b+c+d} \quad (7)$$

$$S_{\text{SOKAL\&SNEATH-II}}=\frac{2(a+d)}{2a+b+c+2d} \quad (8)$$

$$S_{\text{ROGER\&TANIMOTO}}=\frac{a+d}{a+2(b+c)+d} \quad (9)$$

$$S_{\text{FAITH}}=\frac{a+0.5d}{a+b+c+d} \quad (10)$$

$$S_{\text{GOWER\&LEGENDRE}}=\frac{a+d}{a+0.5(b+c)+d} \quad (11)$$

$$S_{\text{INTERSECTION}}=a \quad (12)$$

$$S_{\text{INNERPRODUCT}}=a+d \quad (13)$$

$$S_{\text{RUSSELL\&RAO}}=\frac{a}{a+b+c+d} \quad (14)$$

$$D_{\text{HAMMING}}=b+c \quad (15)$$

$$D_{\text{EUCLID}}=\sqrt{b+c} \quad (16)$$

$$D_{\text{SQUARED-EUCLID}}=\sqrt{(b+c)^2} \quad (17)$$

$$D_{\text{CANBERRA}}=(b+c)^{\frac{2}{2}} \quad (18)$$

$$D_{\text{MANHATTAN}}=b+c \quad (19)$$

$$D_{\text{MEAN-MANHATTAN}}=\frac{b+c}{a+b+c+d} \quad (20)$$

$$D_{\text{CITYBLOCK}}=b+c \quad (21)$$

$$D_{\text{MINKOWSKI}}=(b+c)^{\frac{1}{1}} \quad (22)$$

$$D_{\text{VARI}}=\frac{(b+c)}{4(a+b+c+d)} \quad (23)$$

$$D_{\text{SIZEDIFFERENCE}}=\frac{(b+c)^2}{(a+b+c+d)^2} \quad (24)$$

$$D_{\text{SHAPEDIFFERENCE}}=\frac{n(b+c)-(b-c)^2}{(a+b+c+d)^2} \quad (25)$$

$$D_{\text{PATTERNDIFFERENCE}}=\frac{4bc}{(a+b+c+d)^2} \quad (26)$$

$$D_{\text{LANCE\&WILLIAMS}}=\frac{b+c}{(2a+b+c)} \quad (27)$$

$$D_{\text{BRAY\&CURTIS}}=\frac{b+c}{(2a+b+c)} \quad (28)$$

$$D_{\text{HELLINGER}}=2\sqrt{\left(1-\frac{a}{\sqrt{(a+b)(a+c)}}\right)} \quad (29)$$

$$D_{\text{CHORD}}=\sqrt{2\left(1-\frac{a}{\sqrt{(a+b)(a+c)}}\right)} \quad (30)$$

$$S_{\text{COSINE}}=\frac{a}{\sqrt{(a+b)(a+c)}^2} \quad (31)$$

$$S_{\text{GILBERT\&WELLS}}=\log a-\log n-\log\left(\frac{a+b}{n}\right)-\log\left(\frac{a+c}{n}\right) \quad (32)$$

$$S_{\text{OCHIAI-I}}=\frac{a}{\sqrt{(a+b)(a+c)}} \quad (33)$$

$$S_{\text{FORBESI}}=\frac{na}{(a+b)(a+c)} \quad (34)$$

$$S_{\text{FOSSUM}}=\frac{n(a-0.5)^2}{(a+b)(a+c)} \quad (35)$$

$$S_{\text{SORGENFREI}}=\frac{a^2}{(a+b)(a+c)} \quad (36)$$

$$S_{\text{MOUNTFORD}}=\frac{a}{0.5(ab+ac)+bc} \quad (37)$$

$$S_{\text{OTSUKA}}=\frac{a}{((a+b)(a+c))^{0.5}} \quad (38)$$

$$S_{\text{MCCONNAUGHEY}}=\frac{a^2-bc}{(a+b)(a+c)} \quad (39)$$

$$S_{\text{TARWID}}=\frac{na-(a+b)(a+c)}{na+(a+b)(a+c)} \quad (40)$$

$$S_{\text{KULCZYNSKI-II}}=\frac{\frac{a}{2}(2a+b+c)}{(a+b)(a+c)} \quad (41)$$

$$S_{\text{DRIVER\&KROEBER}}=\frac{a}{2}\left(\frac{1}{a+b}+\frac{1}{a+c}\right) \quad (42)$$

$$S_{\text{JOHNSON}}=\frac{a}{a+b}+\frac{a}{a+c} \quad (43)$$

$$S_{\text{DENNIS}}=\frac{ad-bc}{\sqrt{n(a+b)(a+c)}} \quad (44)$$

$$S_{\text{SIMPSON}}=\frac{a}{\min(a+b,a+c)} \quad (45)$$

$$S_{\text{BRAUN\&BANQUET}}=\frac{a}{\max(a+b,a+c)} \quad (46)$$

$$S_{\text{FAGER\&MCGOWAN}}=\frac{a}{\sqrt{(a+b)(a+c)}}-\frac{\max(a+b,a+c)}{2} \quad (47)$$

$$S_{\text{FORBES-II}}=\frac{na-(a+b)(a+c)}{n\min(a+b,a+c)-(a+b)(a+c)} \quad (48)$$

$$S_{\text{SOKAL\&SNEATH-IV}}=\frac{\frac{a}{(a+b)}+\frac{a}{(a+c)}+\frac{d}{(b+d)}+\frac{d}{(b+d)}}{4} \quad (49)$$

$$S_{\text{GOWER}}=\frac{a+d}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (50)$$

$$S_{\text{PEARSON-I}}=\chi^2 \text{ where } \chi^2=\frac{n(ad-bc)^2}{(a+b)(a+c)(c+d)(b+d)} \quad (51)$$

$$S_{\text{PEARSON-II}}=\left(\frac{\chi^2}{n+\chi^2}\right)^{1/2} \quad (52)$$

$$S_{\text{PEARSON-III}}=\left(\frac{\rho}{n+\rho}\right)^{1/2} \text{ where } \rho=\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (53)$$

$$S_{\text{PEARSON\&HERON-I}}=\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (54)$$

$$S_{\text{PEARSON\&HERON-II}}=Cos\left(\frac{\pi\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}\right) \quad (55)$$

$$S_{\text{SOKAL\&SNEATH-III}}=\frac{a+d}{b+c} \quad (56)$$

$$S_{\text{SOKAL\&SNEATH-V}}=\frac{ad}{(a+b)(a+c)(b+d)(c+d)^{0.5}} \quad (57)$$

$$S_{\text{COLE}}=\frac{\sqrt{2}(ad-bc)}{\sqrt{(ad-bc)^2-(a+b)(a+c)(b+d)(c+d)}} \quad (58)$$

$$S_{\text{STILES}}=\log_{10}\frac{n(|ad-bc|-\frac{n}{2})^2}{(a+b)(a+c)(b+d)(c+d)} \quad (59)$$

$$S_{\text{OCHIAI-II}}=\frac{ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (60)$$

$$S_{\text{YULEQ}}=\frac{ad-bc}{ad+bc} \quad (61)$$

$$D_{\text{YULEQ}}=\frac{2bc}{ad+bc} \quad (62)$$

$$S_{\text{YULEw}}=\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}} \quad (63)$$

$$S_{\text{KULCZYNSKI-I}}=\frac{a}{b+c} \quad (64)$$

$$S_{\text{TANIMOTO}}=\frac{a}{(a+b)+(a+c)-a} \quad (65)$$

$$S_{\text{DISPERSON}}=\frac{ad-bc}{(a+b+c+d)^2} \quad (66)$$

$$S_{\text{HAMANN}}=\frac{(a+d)-(b+c)}{a+b+c+d} \quad (67)$$

$$S_{\text{MICHAEL}}=\frac{4(ad-bc)}{(a+d)^2+(b+c)^2} \quad (68)$$

$$S_{\text{GOODMAN\&KRUSKAL}}=\frac{\sigma-\sigma'}{2n-\sigma'} \text{ where } \sigma=\max(a,b)+\max(c,d)+\max(a,c)+\max(b,d),\ \sigma'=\max(a+c,b+d)+\max(a+b,c+d) \quad (69)$$

The binary feature vector is one of the most common representations of patterns and measuring similarity and distance measures play a critical role in many problems such as clustering, classification, etc

Table 1 OTUs Expression of Binary Instances i and j

j \ i	1 (Presence)	0 (Absence)	Sum
1 (Presence)	$a=i\bullet j$	$b=\bar{i}\bullet j$	$a+b$
0 (Absence)	$c=i\bullet \bar{j}$	$d=\bar{i}\bullet \bar{j}$	$c+d$
Sum	$a+c$	$b+d$	$n=a+b+c+d$

$$S_{\text{ANDERBERG}}=\frac{\sigma-\sigma'}{2n} \quad (70)$$

$$S_{\text{BARONI-URBANI\&BUSER-I}}=\frac{\sqrt{ad}+a}{\sqrt{ad}+a+b+c} \quad (71)$$

$$S_{\text{BARONI-URBANI\&BUSER-II}}=\frac{\sqrt{ad}+a-(b+c)}{\sqrt{ad}+a+b+c} \quad (72)$$

$$S_{\text{PEIRCE}}=\frac{ab+bc}{ab+2bc+cd} \quad (73)$$

$$S_{\text{EYRAUD}}=\frac{n^2(na-(a+b)(a+c))}{(a+b)(a+c)(b+d)(c+d)} \quad (74)$$

$$S_{\text{TARANTULA}}=\frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}=\frac{a(c+d)}{c(a+b)} \quad (75)$$

$$S_{\text{AMPLE}}=\left|\frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}\right|=\left|\frac{a(c+d)}{c(a+b)}\right| \quad (76)$$

A Survey of Binary Similarity and Distance Measures
Seung-Seok Choi, Sung-Hyuk Cha, Charles C. Tappert
Department of Computer Science, Pace University
New York, US

研究方法設計 - 相關研究法

感謝聆聽、敬請指教!

顧志遠 教授

2023年12月5日

中原大學之建校，本基督愛世之忱，
以信、以望、以愛，致力於中國之高等教育，
旨在追求真知力行，以傳啓文化、服務人類。