

緒論

信度是測驗的主要特徵之一。信度係指標準測驗情境下，於不同時間，以相同或複本測驗測量同一群受測者所得分數的一致性或穩定性。為能探究測驗信度，古典測驗理論假設受測者 p 的觀測分數（ X_p ）為其真實分數（ τ_p ）與誤差分數（ E_p ）的線性組合，即，

$$X_p = \tau_p + E_p \quad (1)$$

其中，真實分數 τ_p 為受測者 p 在標準測驗情境下，以相同或複本測驗予以測試無限多次後所得分數的平均值，即， $\varepsilon(X_p) = \tau_p$ 。顯然，參數 τ_p 與隨機變數 E_p 無法觀測， $\varepsilon(E_p) = 0$ ，且 $\sigma_{X_p}^2 = \sigma_{E_p}^2$ 。

就一群受測者而言，他們的觀測分數（ X ）、真實分數（ T ）與誤差分數（ E ）隨個人真實能力不同、受測時身心狀況、測驗情境與試題取樣等因素隨機變動，因此， X 、 T 與 E 皆為隨機變數。將式（1）的觀念稍做推展，古典測驗理論假設

$$X = T + E \quad (2)$$

由於 $\varepsilon(E_p) = 0$ ，對所有不同能力受測者都成立，因此， $\varepsilon(E|T) = 0$ ；即， E 對於 T 的迴歸斜率係數 $\beta = (\sigma_E / \sigma_T)\rho_{ET} = 0$ 。通常， σ_E 與 σ_T 不會為零，因此， $\rho_{ET} = 0$ ，誤差分數與真實分數零相關，而且，

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (3)$$

依據式（3），一群受測者觀測分數的變異源自：這群受測者間真實分數的差異，與測量真實分數時所產生的誤差變異。變異數 σ_T^2 反映這群受測者間真實分數的差異， σ_E^2 則反映測量真實分數時所產生的誤差變異；因此，古典測驗理論學者很自然地定義一測驗分數 X 的信度為：

$$\rho_X^2 = 1 - \frac{\sigma_E^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_X^2} \quad (4)$$

母體信度 ρ_X^2 需要估計，對於測驗評分不受評分者主觀判斷的客

觀測驗，文獻上常被論及用以估計測驗信度的方法包括：再測、複本與內部一致性方法。再測與複本方法必需施測兩次或使用兩份平行測驗，實務上不易進行；內部一致性方法則僅根據一次測驗的結果來估計母體信度。內部一致性方法的基本觀念是將一個完整測驗分解為兩個或多個子測驗，這些不同子測驗中任意成對子測驗觀測分數的共變異數反映受測者真實能力差異的部分比例，而所有成對子測驗的共變異數總和則反映完整測驗中受測者真實能力差異的變異。因此，將此共變異數總和除以總測驗變異即得信度估計值。折半信度（Flanagan, 1937）、KR20（Kuder & Richardson, 1937）與 α 係數（Cronbach, 1951）等都屬於內部一致性方法估計得到的信度。

折半方法係將完整測驗折半分解為兩個等長的子測驗，然後再以這兩個子測驗觀測總分相關係數的Spearman-Brown（1910）校正數值估計測驗信度。一般教科書上提議的折半分解方法包括：隨機分解、奇—偶數題分解與先將題目依難度排序，再做奇—偶題分解等三種方法。直觀的想法，第三種折半方法較有可能產生近似複本的子測驗，但隨機分解並未受到排斥。

在假設兩折半測驗 Y_1 與 Y_2 本質上 τ -等價的情況下，Flanagan（1937）定義完整測驗 X 的信度為：

$${}_F\rho_X^2 = 2\left(1 - \frac{\sigma_{Y_1}^2 + \sigma_{Y_2}^2}{\sigma_X^2}\right) \quad (5)$$

不同於Flanagan的折半考慮方式，Cronbach（1951）將完整測驗中的 n 個項目視為 n 個子測驗 $\{Y_1, Y_2, \dots, Y_n\}$ ，並且在假設這 n 個子測驗本質上 τ -等價的情況下，定義 α 係數為完整測驗的信度：