

# 信度與效度

信度( Reliability )與效度( Validity)是所有測量上常用的指標，可是信度和效度的名稱也常令人混淆和困擾，所以我們應該把它弄清楚。

## 信度

「信度」是指試驗結果的一致性、穩定性及可靠性，一般多是以內部一致性來表示信度的高低。信度高即表示該試驗的結果愈一致、穩定與可靠。系統誤差對信度沒什麼影響，因為系統誤差總是以相同的方式影響測量值的，因此不會造成不一致性。反之，隨機誤差可能導致不一致性，致使信度降低。

一般通過使用同一量表(問卷)進行不同調查，然後分析其結果之間關聯的方法來評價信度。如果關聯密切，就表示各測量結果具有一致性，則認為量表是可信的。

信度主要有：再測信度、內部一致性……等：

### ● 再測信度

用同樣的量表(問卷)，對同一組受訪者在盡可能相同的情況下、不同的時間(一般相距 2 到 4 周)進行兩次測量，以兩次測量結果的相關分析或差異的顯著性檢驗，來評價此量表的信度，越是相關，差異越是不顯著則信度越高。也就是說，同樣的測量工具，對相同的對象測了二次，若二次間差得很多，是不是表示測量的工具可信度不高？

用再次測量法評價信度也存在以下問題：

1. 結果的關係與時間間隔密切相關，時間間隔越長，信度一般就會越低。
2. 最初的測量可能會影響被測特徵。例如，測量人們對高脂肪食品的態度，可能會喚起他們更為關心健康問題，從而對低脂食品持更肯定的態度。
3. 要實施重覆測量實務上有時是很困難的，例如測量消費者對某種新上市產品的反應。

4. 第一次測量的答案可能會對以後測量有影響。受訪者可能會試圖回憶第一次受測時給出的答案，並有意的再給相同回答。
5. 在兩次測量之間若剛好發生一個有利或不利的訊息，可能會使受訪者的態度跟著轉變。
6. 再測信度的相關係數可能會因被測項目自身之間的相關而偏高。兩次測量中，同一項目自身之間的相關性要比不同項目間的相關性高。因此，即使不同項目之間的相關性很低，也可能得到較高的再測相關係數。

- 內部一致信度

內部一致信度主要用於評價累加量表的信度。在累加量表中，是將各個測量項目的得分加在一起來得到一個總分。每個項目要測量測量的是整個量表的某個方面，而各項目之間的特徵而言應該是一致的。內部一致信度強調的是組成量表的各測量項目內部之間的一致性。

「折半信度」是測量內部一致性的簡單方法。量表中的項目被分成兩半並計算測量結果的相關係數，如果這兩半的結果相關係數高，說明量表內部的一致性高。量表的項目可按序號的單、雙分為兩半，也可以隨機的方式分，有興趣的讀者可進一步參考相關書籍。

## 效度

「效度」表示一項研究的真實性和準確性程度，也就是有效的程度。它與研究的目標密切相關，一項研究所得結果必須符合其目標才是有效的，因而效度也就是達到目標的程度。當然效度也是相對的，是針對特定目標而言的，因此只是程度上的差別。

在測量方面，效度指一種測量手段能夠測得預期結果的程度。效度與信度的關係為：信度是效度的必要條件，但不是充分條件。也就是說一個測量的效度要高，其信度必須高，而一個測量的信度高時，效度並不一定高。

效度主要有：表面效度、內容效度、建構效度和效標關聯效度……等：

- 表面效度

表面效度是指從表面上來看，測量工具是否明顯的、直接的與測量的目的有關。可是表面效度高有效的程度不一定高，同樣的表面效度低有效的程度也不一定低，例如大家都有經驗，我們在監理處考駕照時，有紅燈起步、路邊停車、S型彎道、直線加速……等考試項目表面上看起來是符合表面效度的，但是真正上路又是另一回事，所以表面效度不能算是一種真正的效度指標。

- 內容效度

內容效度是指測驗題目對有關內容或行為範圍取樣的適當性。成就測驗和熟練測驗特別注重這種效度。例如，在學校考試的成就測驗中，表面效度就是指考試題目是不是根據教學大綱和教材內容中出的，內容效度就是判斷測驗題目（內容）是否是「適當」，符合它欲測的目標。由於這種衡量效度的方法必須針對課程的目標和內容，以系統的邏輯方法詳細分析題目的內涵，故又稱課程效度或邏輯效度。

確定內容效度的方法主要有兩種：

1. 專家判斷，即由專家判斷測驗題目與測驗目的的符合性；
2. 統計分析，即以一組受試者在取自同樣內容範圍的兩個獨立測驗上得分的相關作出評估。

內容效度很容易與表面效度相混淆，實際上兩者意義不同。表面效度是指從外表（如測驗的材料及用語、試題的印刷等）直觀地看，測驗題目與測量目標的一致程度，它與內容效度所指測驗在實際測量上的有效程度不同。從技術意義上嚴格地說，表面效度並不是效度，但為了取得受試者的信任與合作，表面效度也不可忽視。

- 建構效度(Construct Validity)

建構效度或稱構念效度，是指測驗能測量理論的概念或特質之程度。它主要適用於心理測驗，其目的是以心理學的概念說明和分析測驗分數的意義，即從心理學的理論觀點對測驗的結果加以解釋和探討。在心理學上，所謂構想是指心理學理論所涉及的抽象而屬假設性的概念、特質或變數，如智力、焦慮、機械能力傾向、成就動機等，通常採用某種操作性定義並用測驗來測量。

- 效標關聯效度

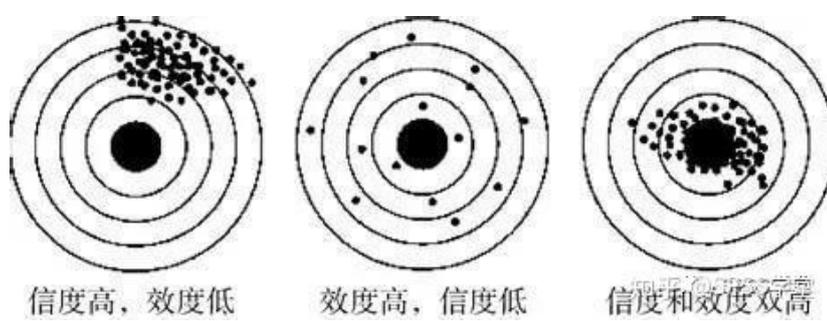
指測驗分數與效度標準的一致程度。效度標準簡稱效標，是足以反映測驗所欲測量或預測的特質的獨立量數，並作為估計效度的參照標準。測驗分數與效標的一致程度以二者的相關係數表示，係數越大測驗的效度越高。效標關聯效度可分同時效度和預測效度。同時效度指測驗分數與當前的效標之間的相關程度，例如入學考試成績與在校學業成績的關係。預測效度指測驗分數與將來的效標之間的相關程度，例如畢業成績與進入社會後的職場表現之間的相關程度。此外，它對人員的甄選、分類與安置工作等也甚為有用。

測量的各種效度區別在於各自強調的方面不同。一個測驗可以有各種效度，每種效度視使用者的具體目的而定，因此，一般不存在測驗的統一效度。但各種效度又是相互聯繫和補充的。

## ● 信度和效度的關係

信度是效度的基礎。測量要有效度必需有信度，沒有信度就沒有效度。一項測量如果信度低，則效度也低，因為若測量不能穩定地測量所需測量的對象，就不能有效的說明測量對象；如果信度高，效度可能高也可能低。一項測量如果效度高，則信度一定高；但是，如果效度低，信度可能高也可能低。值得注意的是，信度與效度二者缺一測量就等於沒有用。

我們可以打靶的概念來了解效度與信度之間的關係。測量目標就是射中靶心，信度就是彈痕的分佈形態。



附錄：請以信度效度的觀點，討論以下個案：



- 某市清潔大隊招考儲備清潔隊員考試，吸引了高達 5,600 多人報名；第一階段筆試後，共有 1,900 多名考生進入第二階段體能測驗（身障組則為口試）。體能測驗的規則為：男性扛 15 公斤沙包，女性扛 8 公斤沙包，進行折返跑速度競賽，折返距離共 60 公尺，依時間快慢給分數。

環保局長表示，體測成績將會轉換為分數；考量男女先天生理差異，並分析過去考試體測成績，再參考專家學者建議及其他縣市作法後，訂出男女不同的計分量表。另外，為維護考試公平性，本次體能測驗特地架設紅外線終端攝影設備並全程錄影，試後即可得知自己的成績。

本次考試預計錄取 300 名儲備清潔隊員，包含一般組 270 名、原民組 20 名及身障組 10 名，將視各區中隊隊員退（離）職狀況，按成績高低依序遞補進用，估計將在 3 年內全數進用完畢。此次清潔隊員的遴選過程完全公平、公正、公開，期勉這些準清潔隊員們未來加入大隊時，能保持考試時的拼勁，創造更美好的市容。