

103 年公務人員特種考試身心障礙人員考試試題

考試別：身心障礙人員考試

等 別：四等考試

類 科：教育行政

科 目：教育測驗與統計概要

一、試以單一樣本 t 檢定為例，回答下列問題：

(一)寫出檢定統計量之公式。

(二)寫出影響統計檢定力之因素。

(三)說明第一類型錯誤率、第二類型錯誤率、檢定力三者之關係。

【擬答】：

$$t = \frac{\bar{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}}$$

(一)統計檢定力 (Power) 是指正確拒絕虛無假設的機率，而影響統計檢定力的主要因素為：

1. 顯著水準，水準愈低統計檢定力愈低。
2. 樣本大小，樣本愈大統計檢定力愈高。
3. 母群變異量，變異量愈大統計檢定力愈低。
4. 效果值大小，差異量愈大統計檢定力愈高。
5. 測驗工具之信、效度，信效度愈好統計檢定力愈高。

(二)由於犯第一類型錯誤率所導致的後果通常較為嚴重，所以第一類型錯誤率 α 常會設得較低，此時研究者會付出第二類型錯誤率 β 較高的代價，也就是二者錯誤的機率間存在抵換 (trade-off) 效果；而同時降低兩種錯誤機率的唯一方法為提高樣本量的使用，但此一方法又常受限於個別研究所面對的研究限制；因而，研究者只得在既定的第一類型錯誤率下，尋覓犯最低第二類型錯誤率的最佳檢定方法 $(1 - \beta)$ 。

二、在計算積差相關係數或建立迴歸模型之前，通常先要繪製散佈圖 (scatter plot)，何謂散佈圖？繪製散佈圖之目的為何？

【擬答】：

(一)探討雙變項資料關係程度，將觀察值在二變項之測量值，標示於平面座標 X 軸和 Y 軸上，藉以檢測全體觀察值資料在二個變項上的分配型態。(王保進，民 95)

(二)散佈圖之繪製目的：

1. 檢定變項間相關強弱程度與方向。
2. 檢定是否具有曲線相關。
3. 檢視資料中是否出現極端值。
4. 檢視二變項相關是否為線性。

三、一個良好的測驗，除了要看其信效度資料外，對其常模之建立也應詳加判斷。某研究者明年將為臺北市高一學生編製一份國文科成就測驗，請您從良好常模應具備特性的角度提出具體建議，讓他在建立常模前可以參考。

【擬答】：

(一)新近性：測驗的常模往往在社會快速變遷之下容易過時，不適合在用來解釋測驗分數，這些改變對成就測驗的影響最大，測驗發展者需要視情形不斷地重新修訂常模，測驗使用者在使用時也應注意常模建立的時間。故其常模之建立須考量編製測驗的時間點。

(二)代表性：常模樣本的各项特徵是否能代表其所欲推論的群體，即樣本特徵與母群吻合的程度。故應採隨機抽樣降低抽樣誤差，以增進常模的代表性。

(三)適切性：即關連性，在使用常模解釋分數時，必須詳查常模樣本的性質是否和擬做測驗的對象相類似，或依據這種常模表所做決定是否適當。故常模樣本性質須和測驗對象(臺北市

公職王歷屆試題 (103 身心障礙特考)

高一學生)相符。

(四)樣本大小：常模樣本人數若太小，通常不具足夠的穩定度，容易出現誤差，也就是若再抽一群人數相同的樣本出來施測，前後二次測驗分數的分佈狀況容易產生不同的情況，所以常模的樣本人數不宜太少。故在具有代表性之下，樣本人數愈多愈好。

四、某教授教育統計學期中考分成兩部分出題，第一部分為是非與選擇題，第二部分為綜合題。第一部分的分數以 X 表示，第二部分的分數以 Y 表示。針對班上 56 位同學施測，結果算得 $S_X=7$ ， $S_Y=10$ ， $r_{XY}=0.80$ ，請回答下面問題：

- (一)計算全班同學期中考總分之變異數。
- (二)求以 X 預測 Y 之未標準化之迴歸係數。
- (三)求以 X 預測 Y 之標準化之迴歸係數，並說明其功能。
- (四)您覺得這位教授出題時需要分別出兩部分嗎？為什麼？

【擬答】：

- (一) $S_{(X+Y)}^2 = S_X^2 + S_Y^2 + 2S_{XY} = S_X^2 + S_Y^2 + 2r_{XY}S_XS_Y = 49 + 100 + 2 \times 0.8 \times 7 \times 10 = 261$
- (二) $b_{Y.X} = r_{XY} \cdot S_Y / S_X = 0.8 \times 10 / 7 = 1.143$
- (三) $\beta = r_{XY} = 0.80$ ，無單位，可進行不同單位自變項之預測比較。
- (四)需要，因 β 值愈大表示該變項預測力愈強，而使得整體迴歸方程式具有顯著意義。

五、級學校經常會使用總結性評量 (summative assessment)，請回答下面問題：

- (一)何謂總結性評量？
- (二)總結性評量之功能為何？
- (三)請分別說明教學過程中常用總結性評量中的「標準化成就測驗」及「最佳作品資料檔案」之優缺點及如何改進？

【擬答】：

- (一)屬於教學效果評量方式之一，它的特點是在教學活動結束之後才實施，目的是為了確定教學目標達成的程度。通常在教學的課程或單元結束後，為了確定教學目標是否有達成，以及學生精熟預期學習結果的程度，並評定學生的成績等級，必須進行總結性的評量。一般來說，學校定期舉行的週考、月考和期考，都可視為總結性評量。此種評量與形成性評量及診斷性評量不同之處在於，它並不偏重於發現學習困難和改進教學，而是較注重在學生成績等級的評定，以及確定學生學習的精熟程度。
- (二)評量的範圍較大，大多是以整個教學活動的單元目標為根據，通常是取能夠代表學習內容的樣本來作為試題，因此可從評量結果中檢討得失，以提供往後教學的改進與參考。因此，雖然總結性評量的主要目的是在評定成績，但是也可作為評鑑課程目標和教學效果是否適當的資料。

(三)1. 「標準化成就測驗」之特徵：

- (1)高品質的試題：題目是由教師及專家共同編擬，並在試題分析後，再從中選擇適當的難度與鑑別度。
- (2)實施與計分的指導說明：使不同使用者有一致的實施程序。
- (3)解釋分數的常模，讓使用者能知道測驗分數在團體中的相對位置。
- (4)可比較的複本測驗：複本測驗，可使讓有所需要的學生，再次接受測驗不被記憶影響，也可了解教學前與教學後的差異。
- (5)測驗編制手冊：做為測驗實施語解釋的指引，或是評鑑品質的依據。

優點是，可減少誤差分數變異提升信度；缺點是，編制費時成本高。故必須先釐清進行評量的目的是什麼，才可以根據此目的來選擇適當的測驗形式，也因為目的的不同，各種評量的使用時機也會因此而不同。

2. 「最佳作品資料檔案」：即歷程檔案評量，也就是學生作品錦囊，以在學習過程中蒐集學生在某學習領域表現的相關作品、文件，做為評量學生學習表現結果的評量方式。歷程檔案評量的內容包括最初構思與最後完成的作品，同時也包含學生的日誌、文章、相片等檔案資料的分析。它不僅是儲存學生學習檔案的資料盒，更提供師生關於學習與個人發展的

公職王歷屆試題 (103 身心障礙特考)

對話入口。在這些檔案資料中，師生皆可知道自己在學習過程中所付出的努力、進步情形和達成學習目標的程度。學生有機會自我檢視、評鑑和反省自己的學習，以成為積極主動的自我評量者。此評量優點在詳實紀錄了學生學習前的經驗態度、學習過程、學習困難與學習成果等等的相關資料。缺點在費時、費力，而且有公平性、評分上規準的問題。為增進客觀性，應強調師生雙向共同參與。

六、某人編製一個團體式智力測驗，並考驗其效標關聯效度，試寫出性質不相同之三種效標及其優缺點。

【擬答】：

效標關聯效度又稱實徵效度(empirical validity)或統計效度(statistical validity)，係以測驗分數和效標(validity criterion)之間的相關係數，表示測驗的效度之高低。效標即可顯示測驗所欲測量或預測的特質之獨立量數，作為檢定效度的參照標準。由於各種測驗所採用的效標，有的是現時可以獲得的資料，有的則需俟諸將來始能蒐集，故效標關聯效度又分為同時效度(concurrent validity)與預測效度(predictive validity)兩種。效標是建立測驗效度的標準。此一標準通常是測驗分數之外的另一組量數，亦即測驗所計畫的某一行為，或代表那些行為的量數。例如，編製學業性向測驗時，可用學業成績為效標；編製機械性向測驗時，可用熟練機工的作業表現為效標。

(一)學業成就：最適合當作智力測驗建立效度時的參照依據，為學生的學業成就(智力測驗又被稱為學業性向測驗)。包括在學成績、成就測驗分數、畢業成績、教育程度等，亦可作為多元性向測驗與人格測驗之效標。

(二)特殊化訓練成績：各種專門學校的訓練成績，常用作專業性向測驗的效標；空軍飛行員選擇測驗效度考驗，可採用基本飛行訓練成績作為效標，因此特殊化訓練成績與專門學校學習成績均屬「中間效標」，但此僅為工作成功部分效標；「終極效標」才是理想效標，如教師實際教學績效。

(三)實際工作成績：此為最佳外在效標，如主管評分成績、工作表現，特別適用於特殊性向測驗效度考驗，於智力測驗與人格測驗也有某些程度用途。(郭生玉，民95)

七、編製選擇題時，學者們會建議謹慎使用「以上皆非」與「以上皆是」兩個選項，試分別說明其可使用之時機，及其使用時可能出現的缺點。

【擬答】：

當出題者很難想出足夠的選項時，常用「以上皆是」或「以上皆非」來作最後一個選項。不過，這樣的應用並不恰當，因為，學生只要具備部分知識即可作較正確的猜測。

(一)謹慎使用「以上皆非」：當在考英文拼字、數學計算或歷史事件等要求絕對正確答案的題目時，可以使用「以上皆非」作為最後一個選項；但是在最佳答案題型或是只要求最接近的數值時，則使用「以上皆非」就不太恰當。

(二)儘量避免使用「以上皆是」：在單一答案的選擇題中，使用「以上皆是」會使試題變得更簡單，因為學生只要知道其中某兩個答案是對的，就可推斷「以上皆是」就是正確答案；或者是知道其中某一答案是錯的，就可以刪去「以上皆是」這一選項。此外，它只能用在正確答案題型，絕對不能用在最佳答案題型上，因為有好幾個最佳答案豈不是與題意自相矛盾。

八、一般學生學習結果應包括知識、技能與情意，然而一般人都重視前兩者的學習與評量，試：

(一)說明情意評量之困難處。

(二)寫出兩種性質不相同的情意評量方法及其優缺點。

【擬答】：

(一)1. 反應心向：是指受試者有意或無意地變更其在測驗上的反應，而塑造出他內心中所希望呈現的形象，但此形象並不代表真正的自己。反應心向的特徵是它與測驗內容有關，受試者可從測驗內容判斷應如何變更其反應，通常會受到社會期望或受試之特殊目的影響。

公職王歷屆試題 (103 身心障礙特考)

2. 偽裝答案：偽裝答案在認知測驗和情意測驗都可能發生，但以後者的情況較為普遍和嚴重。在認知測驗中，只能偽裝差的分數，但在情意測驗中，偽裝好或差的分數都有可能。受試者回答測驗的態度是真實或說謊，關鍵取決於測驗的目的與他所覺到的測驗結果之用途。
 3. 信效度低：情意測驗的信度通常都比智力測驗或成就測驗低，而情意測驗的效度不但低，而且不易考驗。
- (二) 1. 投射技術之性質：投射測驗之基本假設為——一個人對結構不明確的情況或模糊刺激之反應，會映現出其人格特性。亦即，模糊刺激本身不含有特定意義，但其反應是具有特殊意義的，此一意義來自於受試對測驗刺激的主觀解釋和想法，在不知不覺中投射出其心理上的需求、個性、情緒、動機、內在衝突等訊息，經由測驗者對這些資料進行系統分析，可推論出受試者的性格。因此，投射測驗是以一組不具明確意義的材料組成，在含意上相當模糊曖昧，受測者難以猜測測驗的目的和意圖。投射測驗的發展與解釋和心理分析的原理有關，因此，在施測與解釋上均須由受過臨床訓練之專業心理人員進行。優點：(1)彈性大，被試不受限制，可以任意作出反應；(2)材料僅為圖片，因此可以對沒有閱讀能力的被試進行施測。缺點：(1)評分缺乏客觀標準，測驗的結果難以解釋；(2)對特定行為不能提供較好的預測。如測驗上發現某人有侵犯慾望，但是實際上這個人卻很少出現侵犯行為；(3)需要花費大量的時間。
2. 自陳式量表的性質：自陳式量表是以客觀的程序對人格特質予以評量。其原理主要根據特質論，此理論認為人格結構是由若干種特質組成，個體間在這些特質上有程度上之差異，若能對這些特質加以了解，並設計出適當的評量方式，便可將人格特質量化，並與其它人作比較。大多數自陳式量表採用紙筆測驗，形式為是非題或選擇題。測驗項目的意義明確，且整份量表備有標準化常模，可作為評分和解釋的依據。優點：(1)可操作性強；(2)採用標準化測試的形式；(3)簡單易行，解釋比較容易，可進行自我診斷；(4)客觀、全面，應用非常廣泛。缺點：(1)穩定性差，由於個人的行為隨時間而有所改變，所以個性測試所測量的行為比能力測試的穩定性差；(2)被測試者容易弄虛作假，測試中的問題明顯，稍有頭腦的受試者往往可以不費吹灰之力就能使自己看起來非常適合於某項工作。(3)大多數問卷調查表容易被鑽空子，所以預測效度不太理想。在錄用考核(或入學許可)中進行該類測試時，被測試者往往偏向好的一面，即選擇社會所期望的答案，或把自己表現得更好的傾向。

王