

97 年公務人員特種考試原住民族考試試題

等 別：三等考試

類 科：教育行政

科 目：教育測驗與統計

一、請簡答以下兩小題：

(一)請各舉一個例子說明常模參照測驗 (norm-referenced test) 與標準參照測驗 (criterion-referenced test) 在測驗結果解釋上的差異。

(二)請問百分等級和標準分數適用於常模參照測驗或是標準參照測驗？為什麼？

【擬答】：

(一)1. 常模參照測驗 (Norm-referenced Test, 簡稱 NRT)：

指測驗結果分數並無意義，必須根據分數在團體中位置而加以解釋，採用相對性標準比較，一般以平均水準為參照點，其目的在競爭區辨學生程度高低。如大學聯考、國家高普考試、標準化成就測驗、智力測驗、性向測驗等。

2. 標準參照測驗 (效標參照測驗) (Criterion-referenced Test, 簡稱 CRT)：

根據教學前事先所訂定的絕對性標準加以解釋測驗。一般均以學生所學習知識或技能，判定其「及格或不及格」、「精熟或不精熟」、「通過或不通過」，不需參考他人表現來比較。如國小教師自編測驗 (平時考、小考、隨堂測驗)、國家技師執照考試、Bloom 提倡「精熟學習」、中醫師檢定考試、汽車駕照考試、能力本位教育、學力鑑定考試等。

3. 常模參照測驗 (前者) 與標準參照測驗 (後者) 差異

(1)前者的量尺準點為中間、事後決定的，變異性分數變異性愈大愈好，計分方式與解釋為百分等級或標準分數；標準參照測驗的量尺準點二端、事前決定的，變異性分數變異性愈小愈好，計分方式與解釋為二分類數字 (及格與不及格、精熟與不精熟)。

(2)前者的教學用途分班編組 (綜合性測量) (安置性、總結性評量)，學習範圍較廣，每一範圍試題較少 (強調試題的鑑別力) (最能區辨個別差異、競逐高低)，測驗計畫性質使用雙向細目表；後者的教學用途補救教學 (精熟性測量、個體性表現) (形成性、診斷性評量)，學習範圍較窄，每一範圍試題較多 (強調試題在學生學習工作表現) (描述個體表現)，測驗計畫性質使用詳細的教材領域細目表。

(3)前者的評量功能鑑別 (與別人成就表現比較)，信度估計適合傳統統計方法用相關係數 (重測信度、複本信度、內部一致性信度、評分者信度)，效度考驗內容效度、建構效度、效標關聯效度；後者的評量功能檢定 (具體明確成就領域比較)，信度估計適合使用百分比 (百分比一致性、Cohen 的 K 係數)，效度考驗只適合內容效度。

(二)百分等級和標準分數適用於常模參照測驗。因為百分等級與標準分數為團體內常模 (Within-group Norm) 將個人測驗分數在團體中與其他成員比較相對位置，以確定分數高低程度，同學必須競逐高低、區辨程度符合常模參照測驗的特性；不但有一致與清楚的數量意義，且可作各種統計分析。

二、某縣市教育局委請你規劃國小語文科的學力測驗，你預計測驗會包含客觀型式試題 (40題) 及論文型式的短文寫作 (3題)。在一場測驗編製的會議上，一位教師提到：「學力測驗的信度可採用專家審核及統計分析兩種方式，統計分析採再測信度即可。」

(一)從信度的本質評論該位教師的說法是否恰當。

(二)提出兩種你認為最有必要建立或最適當的信度指標，並說明選擇的原因及建立的程序。

(三)針對上述(二)，說明可透過那些途徑提高信度。

【擬答】：

(一)1. 信度是指相同受試者在不同時間，使用相同測驗測量 (或複本測驗測量多次) 或在不同情境下測量，所得結果一致性。藉以反應真實量數程度的一種指標。二次測量結果相當一致，就表示測量分數可靠性高、穩定性高或具有可預測性。

2. 教師的說法並不恰當。

- (1) 學力測驗的信度採用專家審核並不適當。因為專家、資深教師審核題目是否具有代表性與適切性應為內容效度，信度必須使用統計分析方式進行，例如相關係數探討測驗分數的一致性與穩定性。
- (2) 統計分析採再測信度進行亦不恰當。因為認知測驗容易在短時間內改變其潛在特質，不適用重測信度；測驗題目性質會因重測而改變，像推理與思考題目極易因第一次施測時因思考而貫通原則或瞭解解決方法，在第二次時便只需記憶即可；容易受到練習與記憶影響，如果二次施測間隔時間太短，因記憶前次反應情形，造成二次分數變為假性高相關。

(二) 建立或最適當的信度指標

1. 客觀型式試題使用複本同時信度

- (1) 選擇原因：複本信度可以同時施測不受記憶與訓練的影響、不受特殊問題的影響，如受試者本身與環境因素較一致、減少作弊與應付測驗而做訓練的影響。
- (2) 建立程序：當同一種測驗有兩種以上的複本測驗（測驗性質類似在內容、試題格式、長度、難度、鑑別度均一致）時，可將此兩種複本在同一時間分別實施於同一群受試者，再根據此兩份測驗分數計算其相關係數，即得複本同時信度。

2. 論文型式的短文寫作使用評分者信度的等級相關

- (1) 選擇原因：評分結果會受到評分者主觀判斷和意見的影響時，則須採用評分者信度，以估計數位評分者是否評分一致。係指從測驗卷中隨機抽取一些樣本，由不同的評分者或同一個評分者在不同的時間，對每一份試卷評分，在測驗過程中觀察、記錄、評分、計分等各方面相互間的一致性而言，然後根據每份測驗卷評分的兩個分數計算其相關係數，即得評分者信度。
- (2) 建立程序：
哥爾頓 (F. Galton) 以心理學家斯皮爾曼之名字命名之相關統計法，適用於兩變項皆為次序變項之資料。統計學者多半以等級相關一詞代替評分者信度，即表某些作品單獨被一人評定二次（多次）或二人（多人）評定一次評分結果一致性程度，並以相關係數概念表示。論文型式的短文寫作由兩位評分者分別評定等級，各自評定六個級分制，評分者較能區分和說出五級分和六級分作文的差別，由於評閱委員均受過訓練，兩位評閱委員給的分數相差兩級分以上（含兩級分）的情況很少見，一旦出現則該卷交由核心委員進行複閱，再計算其等級相關。

三、設X和Y兩變項之相關為0.941，又以X預測Y，請根據其描述統計和相關資料回答下列各子題：

	平均數	標準差 (S)	人數
X	129	6.674	25
Y	32	5.527	25

【註】表內標準差為母群標準差之不偏估計值。

- (一) 決定係數是多少？你如何解釋該數值的涵義？請列明計算過程。
- (二) 迴歸預測方程式為何？請列明計算過程。
- (三) 估計標準誤是多少？請列明計算過程。
- (四) 列出其變異數分析摘要表（包括F值）。

【擬答】：

(一) 決定係數 $= r^2 = .941^2 = .8855$ 。決定係數 (Coefficient of Determination) 以 r^2 表示，即表「在效標變項 Y 中，可被預測變項 X 解釋到變異量部分為 88.55%，即 X 變項預測 Y 變項預測正確部分高達 88.55%」。因此在總離均差平方和中，迴歸離均差平方和占其多少百分比，便為判斷預測效果大小重要指標。

(二) 迴歸預測方程式

$$\frac{\hat{Y} - \bar{Y}}{S_y} = r \cdot \frac{X - \bar{X}}{S_x} \quad \therefore \frac{\hat{Y} - 32}{5.527} = .941 \cdot \frac{X - 129}{6.674} \quad \therefore \hat{Y} = .7793X + 132.527$$

公職王歷屆試題 (97 原住民特考)

(三)估計標準誤 $S_{Y.X} = S_Y \sqrt{1-r^2} = 5.527 \sqrt{1-.941^2} = 1.8702$

(四)

變異來源	離均差平方和	自由度	均方	F
迴歸	649.200	1	649.200	177.872
殘差	83.9452	23	3.6498	
總和	733.145	24		

四、某教師想知道編序教學法、啟發式教學法、演講法、及自學輔導法等四種教學方法的教學效果。底下是他的實驗結果，假定此實驗是一個單因子完全隨機的設計，各組的人數皆為7人，顯著水準訂為.05。

	自學法	啟發法	編序法	演講法	總計
n	7	7	7	7	28
\bar{X}_j	22.71	29.86	26.29	23.71	25.64
s_j^2	31.24	15.48	29.90	9.90	

各組間平均數差異的情形如下表所示。

	自學法	演講法	編序法	啟發法
自學法	22.71	23.71	26.29	29.86
演講法	—	1	3.58	7.15
編序法	—	—	2.58	6.15
啟發法	—	—	—	3.57

(提示： $q_{.95(3,24)} = 3.53$ ， $q_{.95(4,24)} = 3.9$)

(一)請根據上述資料，試計算下列變異數分析摘要表內之各項數據，並製作表格填寫於試卷上(需附計算過程)。

Source	SS	df	MS	F
Between				
Within				
Total				

(二)假定變異數分析的結果是虛無假設被拒絕的話，當想用 Tukey 的方法來進行事後比較時，自由度是多少？

(三)若改用 Newman-Keuls 的方法，要考驗 6.15 是否達到顯著時，自由度該是多少？(3分)

(四)根據一般的常識，像上述的兩組之間的平均數差異之事後比較，用 Scheffé 的方法是否比較恰當？請說明理由。

(五)近年來美國心理學會 (APA) 所發行的期刊之主編們一直鼓勵或要求研究者必須在他們所撰寫的研究報告中報導「效果量」(effect size)，在此例中，代表 ANOVA 效果量的 η^2 的值為何？

【擬答】：

$$\text{(一)} SS_{\omega} = (n_i - 1) \sum S_i^2 = (7-1) \times 31.24 + (7-1) \times 15.48 + (7-1) \times 29.9 + (7-1) \times 9.9 = 519.12$$

$$SS_b = \frac{\sum (\sum X^2)}{k} - \frac{(\sum \sum X^2)}{nk} = \frac{(7 \times 22.71)^2 + (7 \times 29.86)^2 + (7 \times 26.29)^2 + (7 \times 23.71)^2}{7}$$

公職王歷屆試題 (97 原住民特考)

$$-\frac{(28 \times 25.64)^2}{28} = 18613.8078 - 18407.4688 = 206.339$$

$$SS_t = SS_b + SS_w = 519.12 + 206.339 = 725.459$$

Source	SS	df	MS	F
Between	206.339	3	68.7797	3.1798
Within	519.12	24	21.63	
Total	725.459	27		

(二) 自由度 = (k, N-K) = (4, 28-4) = (4, 24)

(三) 自由度 = (r, N-K) = (3, 28-4) = (3, 24)

(四) 1. 學者 Scheffé 在 1959 年發明，適用各組人數不同或相同時，或進行複雜比較（每次進行二個以上平均數差異比較），簡稱 S 法。學者 Kirk 在 1982 年提出使用 S 法的統計考驗力會比杜凱法（q 法）的統計考驗力低，因此建議只針對簡單二個平均數比較，使用杜凱法較佳。

2. 因此薛費法犯型 II 錯誤機率大於犯型 I 錯誤的機率，常有研究者在進行薛費法事後比較，將 α 值放大比較容易達到顯著水準。當進行事後比較後，有時會發現 F 考驗值達顯著水準，但多重比較卻沒有一個顯著，尤其在嚴格 Scheffé 法可能發生。

$$(五) \eta^2 = \frac{SS_b}{SS_t} = \frac{206.339}{725.459} = .2844$$

公
職
王