

98 年公務人員、關務人員升官等考試試題

等 別：薦任

類 科：教育行政

科 目：教育測驗與統計

一、解釋名詞或簡答：

(一)標準九 (stanine)

(二)語句完成測驗

(三)Q技術 (Q technique)

(四)百分等級

【擬答】：

(一)標準九分數 (Stanine) 美國空軍於世界第二次大戰時所創造出來的標準分數，屬於非直線轉換。標準九係以每半個標準差為單位，提供一個位數的分數系統，將常態分配曲線的底邊分割成九等份，最右邊 (高分) 為 9 分，最左邊 (低分) 為 1 分，依序占面積為 4%、7%、12%、17%、20%、17%、12%、7%、4%。標準九是常態化標準分數，可以比較不同測驗分數，也可從事數學算術運算。

公式： $Stanine = 2Z + 5$

(二)「語句完成測驗 (Incomplete sentence blank)」屬於人格測驗中的投射測驗，提供一些不完全敘述句子或句根 (Sentence stems) 的刺激，讓受試者完成自由填寫，受試者可以有任何型式答案。如語句完成測驗即為此技術。編製原理與技術與主題統覺測驗、羅夏克墨漬測驗相同，用於建立個人人格投射原理，但結構性則較兩者高，界定於半結構性的，開放式作答方式仍可視為受試者投射出想法、慾望、情緒、價值、恐懼與衝突等。

(三)學者史蒂芬森 (Stephenson) 在西元 1953 年創用，評量者提供大量文字敘述句卡片，要求受試者依自己喜好、重要同意程度，將卡片做若干等級分類，卡片約在 60~90 張之間等級約 9~11 等級。每個等級可依常態分配次數或標準九 (分成九等份) 比例安排，要求受試者依比率放置卡片。Q 技術屬於自陳量表的技術。

(四)屬於團體性常模，表示一個人的分數在某一團體中所占的地位，亦即團體中按某一屬性依序排列且分成一百個等級情況下，某一個人分數能勝過多少個百分等級，以 PR 表示。例如：某一個人得分 70 分，在團體中 100 人勝過 80 人，則表示 $PR = 80$ ， $P_{80} = 70$ (原始分數 70 分便相當於第 80 個百分位數)。百分等級為次序變項。計算公式

$$1. \text{未歸類資料 } PR = 100 - \frac{100R - 50}{N}$$
$$2. \text{分組資料: } PR = \frac{[(P_p - L) \times f_p + F]}{h} \times 100$$

二、從事效標關聯效度的考驗時，適當的效標有何特性？常用之效標有那幾種？

【擬答】：

(一)涵義

1. 效標關聯效度是用經驗性 (實證分析) 方法，研究測驗分數與外在效標關係，探討效度類推性 (某個情境結論能夠推論或應用其他情境證據) 又稱實證效度、經驗效度或統計效度；測驗成績對目前或未來某一行為表現預測力高低。效標關聯效度分為同時效度與預測效度二種，主要的區別方式在測驗使用目的不同與外在效標取得時間的不同。(郭生玉，民 95) 外在效標，指測驗所要預測受試者某些行為或表現的標準，如學業成就、評定成績、實際工作表現。

2. 測驗結果分數與外在效標相關愈高，表示效標效度愈高，亦即測驗分數能有效預測外在效標，亦即選取效標時最重要考量為預測力。(郭生玉，民 95)

(二)適當效標特性

公職王歷屆試題 (98 公務人員、關務人員升官等考試)

效標用來表示測驗所欲測量或預測特質，以做檢定效度的參照標準。適當效標具備四種特性。

1. 適切性(Relevance)

外在效標選擇應與所欲衡量特質相近，亦即充分反映出所欲測量的特質。如以學校總成績做為機械性向測驗效標，其適切性不如採用機械學科成績為效標。優良效標特性首要考量因素。

2. 信賴性(Reliability)

外在效標本身必須具備信賴性、穩定性，否則測驗結果會受情境因素影響產生誤差。且樣本大小必須有一定標準。

3. 避免效標的污染(Contamination)

效標污染即教師在評定受試者時，會受到其他因素影響，如主觀意識、種族、性別、偏見，以致降低效標效度。「效標混淆」現象最有可能發生在效標關聯效度（包括同時效度與預測效度）。

4. 實用性(Availability)

在選擇效標時，效標資料的取得是否容易及方便。關係者可用性高低，如取得效標資料取得費時或成本高，則效標可用性較低。

(三)效標種類 (郭生玉, 民 95)

依測驗的特殊用途，不同效標可用來當效標關聯效度的考驗常用效標有下列數種：

1. 學業成就(Academic Achievement)

最適合拿來當作智力測驗建立效度時的參照依據為學生的學業成就，因為智力測驗又被稱為學業性向測驗。學業成就包括學校成績、成就測驗分數、畢業成績、教育程度、特殊榮譽等；亦可作為多元性向測驗與人格測驗的效標。(郭生玉, 民 95)

2. 特殊化訓練成績(Performance in Specialized Training)

各種專門學校的訓練成績，如學校專門學科能力表現、專門學科學業成績（打字、簿記、醫學、護理、會計）等，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效標，因此特殊化訓練成績與專門學校學習成績均屬於中間效標(Intermediate Criterion)，但此部分僅為工作成功部分效標；終極效標(Ultimate Criterion)才是理想效標，如教師實際教學績效。

3. 實際工作成績(Actual Job-performance)

最佳的外在效標為實際工作成績、主管評分成績、工作表現，特別適用特殊性向測驗效度考驗，在智力測驗與人格測驗也有某些程度用途。

4. 對比團體(Contrasted Groups)

任何可清楚對照二樣本，如音樂性向測驗效度，以音樂資優班和普通班學生在測驗分數比較；如美勞性向測驗效度，以美術班和普通班學生在測驗分數比較。

5. 先前有效測驗(Previously Available Tests)

先前有效測驗可做為新編測驗效標，例如比西量表、魏氏量表，因為有嚴謹理論依據，常被做為新編測驗效標；因此新編測驗可以尋找性質相同且有良好效度先前測驗，可做為建立效度效標。

三、何謂Cronbach α 係數？適用於何種測驗？ α 係數的高低代表什麼意義？

【擬答】：

(一) α 係數

α 係數為內部一致性信度係數，克朗巴賀(Cronbach)在 1951 年在「心理統計學報」發表「測驗的 α 和內在結構」，當測驗採用多重記分方式時傳統庫李信度無法探討其信度，此時必須此使用 α 係數。簡稱阿爾發係數(Coefficient Alpha)。

$$\text{公式： } \alpha = \frac{n}{n-1} \left(1 - \frac{\sum S_i^2}{S_x^2} \right)$$

(二)適用測驗

公職王歷屆試題 (98 公務人員、關務人員升官等考試)

如態度量表李克特式量表(「非常同意」5分,「同意」4分,「無意見」3分,「不同意」2分,「非常不同意」1分)、個別智力測驗的加權計分、行為評定量表,皆採用 α 係數方法。

(三) α 係數的高低代表意義

1. α 係數高低反應測量誤差大小,其誤差來自內容抽樣誤差與抽樣內容異質性的誤差。 α 係數愈高,測驗內容同質性愈高。
2. α 係數與 r_{KR20} 常被稱為內部一致性估計最低限度,因為 α 係數與 r_{KR20} 可以反應題目異質性的誤差;反之若題目同質性二者與折半信度接近。 α 係數為所有折半信度方法的平均數,因此估計內部一致性係數, α 公式優於折半法。
3. α 係數特質:
 - (1)測驗的信度是整分測驗所有題目間彼此正相關函數。
 - (2)亦可適用二分計分法(非對即錯)題目。
 - (3)假設題目間相關相同。
 - (4)所有題目都評量同樣(單一)特質。

四、某人想比較三種教學法在國一數學上的成效,隨機自某校母群中隨機抽取樣本9人,再隨機分派為三組(1.啟發組3人;2.電腦輔助組3人;3.演講組3人)。實驗結果的數學分數如下表,請用獨立樣本單因子ANOVA檢定假設。將(a)至(j)共12項之答案,寫在試卷上(須標題號)。

教學法		
啟發	電腦輔助	演講
6	5	0
7	3	2
5	4	4

(一)自變項: (a), 依變項: (b)

(二)統計假設: H_0 : (c) H_1 : (d)

(三)本研究採何種抽樣分配? (e)

(四)在下列變異數分析摘要表中,將(f)至(j)項之答案分別算出,寫在試卷上(須標題號)。

變異來源	SS	DF	MS	F
教學法(組間)	24	(f)	(g)	(h)
誤差(組內)	12	(i)	(j)	
全體		8		

查表 $F_{.95(2,6)} = 3.46$

(五)比較計算的F值和查表F值,何者較大? (k)

(六)統計假設裁決: (l)

【擬答】:

(一)自變項: (a) 教學法(啟發組、電腦輔助組、演講組); (b) 數學分數。

(二) (c) $H_0: \mu_1 = \mu_2 = \mu_3$; (d) $H_1: \mu_i$ 不全等

(三) (e) F 分配

(四) (f) 2; (i) 6; (g) 12; (j) 2; (h) 6。

(五) $F=6 > F_{.95(2,6)} = 3.46$, 因此F值比查表F值較大。

(六)因為 $F=6 > F_{.95(2,6)} = 3.46$, 落入拒絕區, 拒絕虛無假設, 亦即有充分證據支持三種教學方法對數學分數有顯著差異, 至於到底哪兩組有差異則需進行事後比較。惟犯型一錯誤5%。