

97 年公務人員普通考試

類科：教育行政

科目：教育測驗與統計概要

一、請解釋下列有關測驗的名詞：

- (一)多元智能(multiple intelligence)
- (二)性向測驗(aptitude tests)
- (三)真實性評量(authentic assessment)
- (四)課程本位能力測驗(curriculum-based competency tests)
- (五)工作取樣系統(work sampling system)

【擬答】

- (一)美國心理學家葛敦納 (Gardner) 在 1983 年出版《智力架構：多元智慧理論》(Frames of Mind) 提倡「智力多元論」(Theory of Multiple Intelligence)，強化個體認知的跨文化觀點對人類智慧概念提出革新實用性定義，並放棄傳統以標準化測驗得分定義人類智力，特性包括智能以特定的材料和問題形式顯現在特定領域、每個正常的人都具有八種智能且但強度不同、八種智能在日常生活中是彼此相互影響。強調人類心理能力中包括八種不同智力：語文智力、邏輯—數學智力、空間智力、音樂智力、體能智力、社交智力、自知智力、自然觀察智力。
- (二)性向測驗主要在測量個人學習知識與技能的潛在能力，與預測個體未來發展的可能性。通常包括普通性向（學習一般事務所共同需要的能力）與特殊性向（學習音樂、美術、機械、科學等特殊才能所需的能力）兩種。性向測驗與智力測驗皆為最大表現測驗。
- (三)真實性評量屬於多元評量，多元評量適合「過程知識」(Procedural Knowledge) 的檢驗，較重視思考歷程評量、學生學習精熟進步情形。強調真實情境的評量方式，此一真實情境指在未來情況發生下，可以自我應用的情況，如學習外語能力是否在未來能與外國人交談、學習 CPR 急救術在遇到危機情況是否能應用、學校採取闖關活動來評量學生的學習成就；可以消除課堂中過於理想化教學與日常生活情境脫離。真實評量方式如檔案評量、實作評量、口頭晤談。
- (四)課程本位能力評量是一種整合課程、教學與測驗的一種教學評量模式，屬於一種非標準化測驗 (Nonstandardized Test)。CBA 的特色在於教師在教學過程中，根據實際課程內容所希望達成的形成性目標 (Formative Test)，經常性採取一些內容簡短、施測便捷的測驗，作為告知學生學習是否精熟，以及教師改進下一階段教材教法和補救教學的依據 (Jenkins & Jewell, 1993; Gickling & Thompson, 1985)。強調學科知識的邏輯與作業分析，進一步檢視學生學習精熟程度與錯誤類型，以作為教學策略選擇依據。
- (五)工作取樣系統就是卷宗評量的另一種說法，學者 Paulson and Meyer 在 1991 年提出，卷宗的意義指有目的、系統地蒐集學生在某一領域或數個領域學習、努力與成就的作品。整個卷宗作品蒐集選擇標準、評量標準，都讓學生全程參與，同時內容更包含了學生自我反省的證據 (王文中、吳毓瑩，民 88)。工作樣本 (Work Sample) 需要受試者在實際作業上，親身操作，表現出所要測量的全部真實技能；工作樣本是真實性程度最高的評量方式。

二、在以古典測驗理論(classic testing theory)進行試題分析(item analysis)時，通常要計算每一個個別試題之難度(item difficulty)，與鑑別度(item discrimination)，並檢視試題選項(options)的分布，也就是誘答項分析(distractor/distracter analysis)。

請回答下列有關試題分析的問題：

- (一)請分別說明試題之難度與鑑別度的意義。
- (二)請說明試題之難度與鑑別度的意義。
- (三)何謂誘答項分析？請以四個選項的單選選擇題為例說明之。

【擬答】

- (一)1. 難度：

公職王歷屆試題 (97 高普考)

計算全體受試者答對每個試題的人數占全體總人數的百分比值，這個百分比值稱為「難度指標(數)」(Item Difficulty Index)，使用次序量尺，可以指出題目等級順序或相對難度。P 值愈大，難度愈低；P 值愈小，難度愈高；P 值以 0.5 最恰當(難易適中)。

$$\text{公式：} P = \frac{P_H + P_L}{2} \text{ 或 } P = \frac{R}{N}$$

2. 鑑別度指數：

鑑別度分析的目的，主要在瞭解試題是否具備區別學生能力高低的作用。某個試題鑑別度愈高，表示能明確分辨學生答對與答錯功能很強，亦即高分組會傾向答對，低分組會傾向答錯，試題具備此種辨別作用，即為「試題鑑別度」。一個優良的測驗試題，具有較高鑑別度。

$$\text{公式：二系列相關或 } D = P_H - P_L$$

(二) 難度與鑑別度的關係：

1. 難度是鑑別度的必要條件，鑑別度是難度的充分條件。鑑別度高，難度一定適中，難度不適中，鑑別度一定低；難度適中，鑑別度不一定高，鑑別度低，難度不一定不適中。當鑑別度 0.4 以上、難度為 0.5 時，此份測驗品質較優良。
2. 難度不適中(偏高或偏低)，大多數學生得分為正偏或負偏分配。其樣本變異程度較小，鑑別度一定低，信度低。難度是鑑別度的必要條件，鑑別度是難度的充分條件。鑑別度高，難度一定適中，難度不適中，鑑別度一定低；難度適中，鑑別度不一定高，鑑別度低，難度不一定不適中。當鑑別度 0.4 以上、難度為 0.5 時，此份測驗品質較優良。

(三) 1. 誘答項分析

客觀測驗進行選項誘答力分析目的在進一步提供試題分析的指標，作為判斷測驗試題編製優劣與診斷學生反應組型的依據。分析選擇題是否具有誘答功能，只要分析高低組學生在每個試題選答的次數分配加以判斷即可。國內郭生玉教授提出判斷原則：(1) 每一錯誤選項至少有一低分組學生選擇。(2) 選擇錯誤選項的低分組學生人數應比高分組學生還多。

2. 舉例：

	A 選項*	B 選項	C 選項	D 選項	難度	鑑別度
高分組	450	42	0	357	0.48	0.32
低分組	210	202	0	105		

*為正確答案選項

B 選項誘答力佳，能有效鑑別高分組與低分組

C 選項誘答力不佳，因為高低分組均無人選填

D 選項誘答力不佳，設計不良因為高分組答錯人數高於低分組，無法有效鑑別高分組與低分組。

三、某生在第一學期結束後，其所修的五個科目的成績及其全班在此五個科目成績的平均數及標準差分別如下：

科目	某生成績	全班平均數	標準差
國文	85	83	7
英文	82	81	9
數學	75	70	10
物理	73	71	11
化學	80	78	8

(一) 請根據上表分別計算該生五科成績之標準分數，也即 z 分數。

(二) 若該生與全班同學比較，其最強與最弱的科目是那兩個科目？

(三) 平均數及標準差的應用之一是用於計算變異係數(coefficient of variation；簡稱 CV)，也即標準差相對於平均數的比，並以百分比表示；因此，CV 可用來比較單位不同的變項之相對變異情形。請依以上描述，計算上表五個科目的相對變異情形，並由高至低分別列出

各科目之 CV。

【擬答】

$$\text{(一)國文 } Z = \frac{X - \bar{X}}{S} = \frac{85 - 83}{7} = \frac{2}{7} = .2857 ; \text{英文 } Z = \frac{X - \bar{X}}{S} = \frac{82 - 81}{9} = \frac{1}{9} = .1111$$

$$\text{數學 } Z = \frac{X - \bar{X}}{S} = \frac{75 - 70}{10} = \frac{5}{10} = .5 ; \text{物理 } Z = \frac{X - \bar{X}}{S} = \frac{73 - 71}{11} = \frac{2}{11} = .1818$$

$$\text{化學 } Z = \frac{X - \bar{X}}{S} = \frac{80 - 78}{8} = \frac{2}{8} = .25$$

(二)最強的科目為數學(因為 Z 最大);最弱的科目英文(因為 Z 最小)。

(三)五個科目的相對變異情形,並由高至低分別列出各科目之 CV

$$\text{物理 CV} = \frac{S_x}{\bar{X}} \times 100\% = \frac{11}{71} \times 100\% = 15.4929\%$$

$$\text{數學 CV} = \frac{S_x}{\bar{X}} \times 100\% = \frac{10}{70} \times 100\% = 14.2857\%$$

$$\text{英文 CV} = \frac{S_x}{\bar{X}} \times 100\% = \frac{9}{81} \times 100\% = 11.1111\%$$

$$\text{化學 CV} = \frac{S_x}{\bar{X}} \times 100\% = \frac{8}{78} \times 100\% = 10.2564\%$$

$$\text{國文 CV} = \frac{S_x}{\bar{X}} \times 100\% = \frac{7}{83} \times 100\% = 8.4337\%$$

四、皮爾遜積差相關(Pearson's product-moment correlation, 以 r 表示)係最常用的統計分析方法之一。請回答下列有關皮爾遜積差相關的問題:

(一)從雙變項資料()的關係及測量量尺()的觀點,說明使用此種統計分析方法的條件?

(二)解釋積差相關係數(r)時要注意那些事項或限制?

(三)若有一項研究結果指出 IQ 與學業成就的積差相關係數(r)為 0.90,從變異的觀點,要如何解釋 IQ 對成就的貢獻有多少?

(四)因為前項研究結果並非完全正相關,從變異的觀點,還有多少變異無法解釋?

【擬答】

(一)適用 X 與 Y 變項均為連續變項(等距變項或等比變項)統計處理方式。積差相關係數 r 是一種次序變項,並非比率變項或等距變項,無因果關係,只是具有共變關係。

(二)應用相關研究法分析資料,在解釋相關係數之意義時應注意事項:

1. 相關係數的意義須視樣本大小:

當樣本很少時,即使相關係數很高,未必具有意義,相反地,樣本人數很大時,即使相關係數不高,若相關係數值達到統計顯著水準時,可能也會具有意義;當相關係數達到統計顯著性,表示樣本相關係數很大,以致於母群中兩變項沒有相關(ρ)的可能性很低。

2. 相關係數並非百分比概念:

相關係數值並非表示二變項解釋百分比,如 $r = 0.8$ 並不能解釋二變項有 80% 相關。應使用決定係數(r^2)說明某變項是由另一變項解釋或預測比率。解釋相關係數時勿將其視為等距變項或比率變項,如相關係數 0.8 為 0.4 相關 2 倍,或 0.7 與 0.5 二個相關視為與 0.4 與 0.2 差一樣。通常愈接近 $r = \pm 1$ 的相關係數差其重要性高於低相關間差異。

3. 變項間具有相關並非因果關係:

(1)相關係數並沒有嚴格控制其他影響因素,有可能二變項皆是因,二變項皆是果或者一因一果。如需建立因果關係,最好採用實驗設計,設計加以考驗。

(2)相關係數雖然達到顯著水準,但未必具有實用意義,例如智力和學業成績相關 0.6,雖達統計顯著水準,但根據此相關預測學業成績,僅可預測 36% 變異量,預測實用價值不高。因此實用意義應比統計意義更重要。

4. 相關係數與受試者的變異程度有關:

如果受試者在二變項分數變異程度很大,其相關係數也較大;反之,二變項或其中一變項

公職王歷屆試題 (97 高普考)

分數變異程度較小，其相關係數也較小。

(三) 1. 使用決定係數(Coefficient of Determination)解釋，表效標變項 Y 中之總變異百分比有多少可由預測變項 X 來解釋到的變異量百分比，或由 X 變項預測 Y 變項預測正確部份。因此在總離均差平方和中，迴歸離均差平方和佔其多少百分比，便為判斷預測效果大小重要指標。

2. IQ 預測學業成就其 $r^2 = 0.81$ ，即 $r^2 = \frac{SS_{reg}}{SS_t} = 0.81$ 即表「在學業成就成績中，可被 IQ 解釋

到變異量部份為 81%，即 IQ 預測學業成就成績預測正確部份高達 81%」。

(四) 1. 使用疏離係數的平方解釋。疏離係數或離間係數(Coefficient of Alienation)效標變項 Y 之變異量中，不能由預測變數 X 決定或解釋的變異百分比。這個部份若愈大，即表示無法正確預測部分愈高，亦即由預測誤差所造成部份愈大。

2. IQ 預測學業成就其疏離係數的平方 $= 1 - r^2 = 1 - 0.81 = 0.19$ ，即表「在學業成就成績中，無法被 IQ 解釋到變異量部份 19%，即 IQ 無法預測學業成就成績 19%」。

【高明老師小叮嚀】

今年普考真如老師考前所猜考多元評量、相關係數、試題分析、相對地位量數（非常誇張已考四年），且今年考解釋名詞 5 小題因此高考時同學務必留意老師考前大猜啼的啼行，且需注意高考應有可能考解釋名詞。

公
職
王