

100 年特種考試地方政府公務人員考試試題

等 別：四等考試

類 科：教育行政

科 目：教育測驗與統計概要

一、對照說明下列各子題中兩專有名詞意涵和應用上的差異：

(一)估計標準誤 (standard error of estimate) 和測量標準誤 (standard error of measurement)。

(二)輻合效度 (convergent validity) 和區別效度 (discriminant validity)。

【擬答】：

(一)1. 估計標準誤：係根據迴歸線上分數來預測 Y 產生的估計誤差，當進行無數次的預測會產生無數次預測誤差時，這些誤差所形成的標準差即估計標準誤。進行預測時，希望未被解釋變異愈小愈好，即 $SS_{res} = \sum(Y - \hat{Y})^2$ 愈小愈好，即求估計標準誤最小值。估計標準誤應用在預測分數的區間估計與測量標準誤。

$$\text{公式： } S_{y \cdot x} = S_y \sqrt{1 - r^2}$$

2. 測量標準誤：用以估計受試者測驗之真實分數所在範圍的誤差單位與分數穩定性的信度。在標準情境下，使用相同測驗或複本測驗測量一個受試者許多次，其所得分數的平均數即為個人的真實分數。一般來說，會以受試者的真實分數為中心形成常態分配。而每次測驗實得分數與真實分數的差，稱為測量誤差，這許多次測量誤差分配之標準差即為測驗標準誤，又稱為分數的標準誤差。

$$\text{公式： } SE_{meas} = S_x \sqrt{1 - r_{xx}}$$

3. 估計標準誤應用為測量標準誤：心理測驗領域中，估計標準誤被稱為測量標準誤，採用信度係數作為解釋測驗分數正確程度依據指標，其中信度係數與決定係數意義和概念相同。

(二)學者 Campbell 與 Fiske 在 1959 年提出一份良好的構念效度，可以使用多項特質—多項方法分析法，必須符合下列二條件：

1. 聚斂效度 (輻合性效度) (Convergent Validity)：一份測驗分數必須和相同構念其他測量的分數間存在高相關。如美勞性向測驗的分數與在校美勞科目成績有高相關，即為聚斂效度。

2. 區別效度 (辨別效度) (Discrimination Validity)：一份測驗必須和不同構念或潛在特質其他測量的分數間存在低相關。如美勞性向測驗分數與在校閱讀理解測驗成績有低相關甚至零相關，即為區別效度。

3. 如果構念效度令人滿意條件，重測信度 > 聚斂效度 > 區別效度；相同方法測量相同特質與不同方法測量相同特質所得的相關，應該比相同方法測量不同特質與不同方法測量不同特質所得的相關高。

二、小明的父母很重視小明的學習成就，期中考試小明的數學科之校排名百分等級為 55，國文科的百分等級則為 88，其同學小玲的數學科百分等級為 45，國文科百分等級則為 95，小明的父母覺得很安慰，他們說雖然小明國文比小玲差 7 分，但是數學比小玲多 10 分，整體來說還是贏過小玲。

(一)小明父母對分數解釋的說法是對的嗎？為什麼？

(二)如果直接將兩個科目的原始分數加總，然後比較他們兩人的總分，是合適的作法嗎？為什麼？

(三)在同一標準化測驗的兩個成績之比較應注意那些事項？

【擬答】：

(一)不對。

1. 百分等級量尺單位並不相等，是一種次序量尺，靠近常態分配中央原始分數單位小，二端單位大；例如常態分配位於分配中央人數多，一兩分差距其百分等級差異很大，而在

公職王歷屆試題 (100 地方政府特考)

分配兩端因為人數較少，所以即使分數有很大差異，其百分等級變化卻不大。亦即接近中央原始分數差異會因為百分位數的轉換而擴大，但分配二端原始分數差異則會大幅地縮小。(危芷芬，民 88)

2. 百分為數單位大小不一，越靠近分配的兩端，單位的不等性越明顯。所以它只能指出某人在常態樣本下的相對位置，並不能說出兩個百分位數的差距。

(二) 1. 不合適。

2. 原始分數的優點即是以受試者作答情形或答對題數直接作計算，簡單易懂，大部分人都明瞭計分方式；缺點即沒有考量測驗試題具有不同難度的事實，因而計分容易受到測驗難度的影響，而出現不公平、不精確的現象。直接將兩人的兩個科目的原始分數加總比較總分，可能會因為學生在某一科目分數拿取高分，導致測驗原始分數總分較高而被錄取，但未必比較優秀。

(三) 在同一標準化測驗的兩個成績之比較應注意那些事項？(10 分)

1. 如使用標準分數較容易受到極端分數 (Outliers) 的影響，而有扭曲次數分配的形狀之嫌，對個別受試者的表現會有過度放大或過度縮小之影響。
2. 測驗後除對學生解釋分數外，尚須瞭解學生作答時的動機、生理狀況、情緒等，瞭解是否在最佳狀態下作答。同時考慮文化背景、教育經驗、學生性向等因素的影響。
3. 測驗分數應以該測驗所欲測量特質加以解釋，同一標準化測驗的兩個成績不可能測量完全相同特質，例如某一數學測驗僅測量簡單計算技巧，而另一數學測驗則為較高層次推理問題。
4. 任何測驗均會有測量誤差存在，解釋分數時最好考量誤差大小，如採用應以可信賴範圍解釋，如百分位數帶解釋兩個成績，當兩個測驗的帶狀範圍重疊，則表示測驗分數沒有顯著差異。
5. 同一標準化測驗的兩個測驗成績之比較應也補充資料加以佐證，單一測驗的解釋分數並非瞭解學生唯一資料，須參照其他相關測驗，如解釋 IQ 時，可參考動機、興趣、語文程度、在校成績、文化背景等相關測驗，解釋分數才能客觀深入。

三、張老師出了 20 題是非題考學生的英文字彙能力，在閱讀每一單字後，學生即判斷該字的拼法是否有誤。考完後，全班最低分為 6 分，最高分為 20 分。

(一) 請問此測驗分數屬於名義變項、次序變項、等距變項，或是比率變項？並請說明你的理由。

(二) 請問考生在此測驗得分的全距為多少？

(三) 請問此份資料計算全距合宜嗎？請說明你的理由。

(四) 請問此份資料計算中位數合宜嗎？請說明你的理由。

【擬答】：

(一) 1. 測驗分數屬於等距變項。

2. 理由：因為測驗分數不僅能標示事物與事物間不同特質或類別而指定的數字，亦能指出大小次序（等級順序）、多寡、優劣、高低，尚可以數值方式表示二事物差別之大小的變項。

(二) 全距 = $\text{Max} - \text{min} = 20 - 6 = 14$ 。

(三) 不合宜。理由分述如下：

1. 只考量數值中最大值與最小值，未考量數值中其他數值，若二群體全距相等，但群體數值差異卻有可能不同。
2. 易受極端值影響，因此受抽樣變動影響很大，即人數愈多全距一定較大（抽樣個數愈多，全距則易愈大；抽樣個數愈少，全距則易愈小）。

(四) 不合宜。理由分述如下：

1. 中位數次序變項不適合代數演算，本文測驗分數為等距變項應使用平均數較為合宜。
2. 中位數對觀察值敏感度低，不易進行有母數統計推論。

四、某心理學家欲進行一項研究，比較控制組與實驗組的平均數（依序為 \bar{X}, \bar{Y} ）有無差異，假設每組各有 25 人，且觀察值近似常態分配，母群標準差 $\sigma = 5$ 。

(一) $\bar{X} - \bar{Y}$ 的標準誤為何？

(二) 已知 $\alpha = .05$ ，試問 $H_0: \mu_X = \mu_Y$ vs. $H_0: \mu_X \geq \mu_Y$ 的拒絕區為何？

(三) 如果 $\mu_Y = \mu_X + 1$ ，統計考驗力為何？

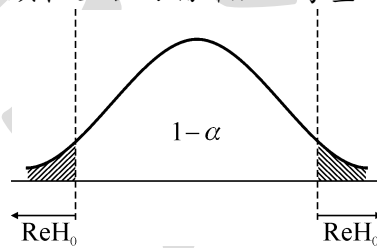
(四) 假設上述考驗的 p 值為 .07，試問該考驗是否會拒絕於 $\alpha = .10$ 的顯著水準？為什麼？

【擬答】：

(一) $\bar{X} - \bar{Y}$ 的標準誤 $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{5^2}{25} + \frac{5^2}{25}} = \sqrt{2}$ 。

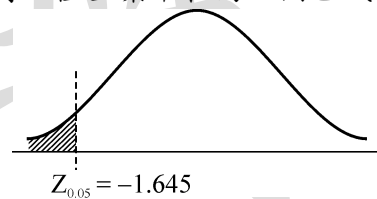
(二) 1. $H_0: \mu_X = \mu_Y$

$H_1: \mu_X \neq \mu_Y$ 此時考慮臨界區必須在左右兩端都加以考量，稱為雙側考驗。



2. $H_0: \mu_X \geq \mu_Y$

$H_1: \mu_X < \mu_Y$ 單側左尾考驗此時 α 值全集中在同一側區域。



(三) 統計考驗力 .8413。

(四) 1. p 值 = .07 < $\alpha = .10$ 表落入拒絕區，有充分證據支持控制組與實驗組的平均數達顯著水準。惟結論說法仍有 0.1 的可能性犯型 I 錯誤。

2. 在某一樣本資料研究結果，去求一個數值為 P，此一 P 值表支持虛無假設的證據與訊息。考驗時可與 α 顯著水準比較，若 P 值 > α ，則接受虛無假設；若 P 值 $\leq \alpha$ ，則拒絕虛無假設，P 值提供客觀不變的證據。其中 P 代表「Probability」亦即「精確機率值」。