

《教育測驗與統計概要》

試題評析

乍看今年的考題，似乎已經告訴未來的應考者，「教育測驗與統計」此一科目的考試，已經慢慢的由重計算演變至重觀念與理論。普考的四個題目中，第一題的信效度的種類與其影響因素在我們講義中的Ch.15、16中皆有清楚列出，而從變異數的角度看信效度的關係則在總複習班補充講義的觀念九更清楚的藉由公式說明二者的關係。第三題考「中央極限定理」在統計假設考驗(檢定)中所扮演的角色，在講義中Ch.9的p.38以及上課中也一直是我們強調的內容。至於第四題考的是綜合的觀念題，只要是班上的同學，絕對對t-test與ANOVA，以及對ANOVA與Regression analysis間的關係瞭若指掌，更別提其所需的假設了！第二題的常模種類在講義Ch.16的p.19有簡單提及，至於其優缺點的比較，則應為此次應考學生唯一較不易掌握所有分數的關鍵。

以一位表現正常的高上同學而言，分數應至少可得75分，更穩健的同學甚至可以取得接近90分的高分。

一、請說明信度 (reliability) 及效度 (validity) 的種類，並分別討論影響信度及效度的因素，以及從變異量的角度來談信度與效度之間的關係。(三十分)

答：(一) 信度的種類

1. 重測驗法: 再測信度。
2. 複本(份)法: 複本(份)信度，又分為
 - a. 等值係數
 - b. 穩定且等值係數
3. 折半法: 折半信度係數，又稱內部一致性係數。
4. 其他信度:
 - a. 庫李信度
 - b. α 係數
 - c. 評分者信度

(二) 效度的種類

1. 內容效度:
2. 效標關聯效度，又分為:
 - a. 同時效度
 - b. 預測效度
3. 構念效度

(三) 影響信度的因素:

1. 測驗的長度
2. 測驗的難度
3. 團體的變異量
4. 估計信度的方法
5. 其他：如測驗之客觀性，測量的機會誤差 ..等。

(四) 影響效度的因素:

1. 測驗的品質
2. 測驗的實施
3. 效標的品質
4. 受試者的個人因素
5. 團體的變異量

(五) 1. 信度的原理:

實得分數 = 真正分數 + 誤差分數

$$X = T + E$$

以變異數的角度來看，

$$s^2_X = s^2_T + s^2_E$$

因此，我們可以界定信度 $r_{XX} = s^2_T / s^2_X = 1 - s^2_E / s^2_X$ ，稱為信度係數，其值介於0與1之間。

2. 效度的原理：

由變異數的角度分析：

$$s_X^2 = s_{CO}^2 + s_{SP}^2 + s_e^2 \quad (\text{等號左右兩邊同除以 } s_X^2), \text{ 得}$$

$$1 = \frac{s_{CO}^2}{s_X^2} + \frac{s_{SP}^2}{s_X^2} + \frac{s_e^2}{s_X^2} \quad \text{其中 } \frac{s_{CO}^2}{s_X^2} \text{ 為效度，因此}$$

$$1 = \text{效度} + \frac{s_{SP}^2}{s_X^2} + \frac{s_e^2}{s_X^2} \quad \text{又 } \frac{s_{CO}^2}{s_X^2} + \frac{s_{SP}^2}{s_X^2} \text{ 為信度，因此}$$

$$1 = \text{信度} + \frac{s_e^2}{s_X^2} \quad \text{所以}$$

$$\text{效度} = \text{信度} - \frac{s_{SP}^2}{s_X^2}$$

二、試說明常模（norms）的種類，並比較各種常模的優缺點。（二十分）

答：（一）發展性常模：年齡，年級，發展商數等

（二）團體內常模：百分等級常模，標準分數等

（三）發展性常模有其使用上之限制，通常使用於在教學上有連續行為的學科，如共同科的英文，國文等；或是針對成長中的學童才有使用與應用之意義。團體內常模雖然看似應用機會較多，但常模的品質對測驗的使用與解釋有很大的影響，因此必須考量常模樣本代表性、樣本規模、以及常模本身的幾個特性，如經濟性、適切性等。

三、請說明什麼是中央極限定理（central limit theorem），並說明此定理在統計假設檢定（例如t檢定、z檢定）上的角色。（二十分）

答：中央極限定理乃有關樣本平均數次數分配之理論。反覆由 (m, s) 之母體中抽取樣本大小為 N 之許多樣本，得許多 \bar{X}_i ，

四、請回答下列問題：（三十分）

（一）請討論t檢定與變異數分析（analysis of variance, ANOVA）這兩種統計方法的目的及思考邏輯之異同。

（二）變異數分析與迴歸分析（regression analysis）二者之間有什麼關係？

（三）請分別說明以上三種統計方法所需要的假設。

答：（一）t-test常用以針對兩組獨立樣本的平均數進行推論統計，特別是在兩組獨立樣本皆為小樣本，母體標準差未知，

但兩母體變異數經由F-test證明無顯著差異時，其標準誤為 $\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ ；而ANOVA又稱F-test，其為對三

個或三個以上的平均數進行顯著性差異時所使用的統計檢定。同樣是對平均數進行檢定，目的皆在找出不同分組間的平均數是否存在差異，以提供選擇較佳方案或決策之參考，ANOVA的思考邏輯利用組內變異與組間變異的比值作為結論的依據。t-test的共用變異數也有類似意義。

（二）ANOVA與Regression analysis之間的關係：

皆透過變異數分析摘要表以及F-test來列出觀測值的變異來源，並做結論。在兩種統計方法中，相同的是，只要組內變異較小(ANOVA)以及殘差變異(Regression analysis)較小，幾乎可以得到拒絕虛無假設的結論；不同的是ANOVA為對三個或三個以上的平均數進行顯著性差異；而Regression analysis在對母體迴歸模式的

預測能力與價值進行結論。

(三) 以上三種統計方法所需的假設為：

1. 變異數相等(同質性)
2. 變異來源可加性
3. 常態母體