

# 《教育測驗與統計概要》

## 試題評析

第一題：本題在本班考前總複習補充閱讀資料與為同學準備的考猜，都精準命中！只要答題時，能夠區分兩種作為主觀評分性強之實作評量的評分方式，並用到本人提供之例子，應可取得高分！答題上沒有任何技巧。

第二題：本題命題以三大類統計量數之計算變化關係的應用為主。臨場若無法順利回答者，只要舉每個變數三個數值為例，進行現場簡單計算推論即可。

第三題：信度與效度的基本題目，只要清楚各種信度與效度的定義即能掌握分數！

第四題：雖然公式計算較多，但都是簡單公式的換算，本題主要測驗考生的耐性與細心。第四小題應該先利用身高數值代入公式，取得體重的估計值後，再進行信賴區間估計。其算式在96普考與97高考都考過，但是公式冗長，不少同學無法順利記憶！在本班總複習講義也曾利用考古題，清楚列出計算式進行詳解！程度好的考生應可以掌握！

總而言之，今年的第二題與第四題考得較細，程度好、觀念清楚的同學應可以取得不錯分數；本班同學在兩題問答題應可取得高分！不過，因錄取名額少，本科分數須在75分以上，對自己才有利！

一、實作評量中，評分設計是發展有效實作評量的重要工作，而教師應依評量目的，選擇適當的評分方法。但不論用何種方法，實作評量常遭遇到教師評分者誤差。(25分)

(一)解釋二種常用的評分方法：整體評分法和分析評分法，並分別舉例說明二種評分法的適用目的。

(二)說明影響實作評量信度和效度的三種教師評分者誤差來源。

**答：**

(一)針對評分常涉及主觀的實作評量，為增進評分效率，避免評分時太過主觀，近來評量專家發展出整體式(holistic rating)法與分析式(analytic rating)法兩類主要的評定方法。其中，整體評分法只對學生表現的品質，參照評分指標或評分規準中對各級分或等第之表現的整體綜合描述，直接給予一個整體性的分數或等第。至於，分析評分法常以一種雙向的表格呈現評分規準的兩個向度。一個向度是一系列用以判斷的規準(criteria)，另外一個向度是三到四個層次的精熟水準(level)，而在這兩個向度所構成的每一個細格裡，都有文字說明以界定每一精熟水準的狀況。視需要，每一個規準也可以依其重要性給予不同加權(weight)。

分析評分法可以應用在國民中學學生寫作測驗、設立音樂評量標準、表演藝術等藝術與人文的實作評量上。因為評分規準是一種等級量尺，它具體地說明了該評量項目的精通程度，用文字詳盡告知學生各自標的層次及要求。以音樂評量為例，Asmus(1999)認為評分規準寫下了學生能夠達成的學習任務，可提供學生未來改善其音樂表現的依據。整體分析法則用一個等第評定學生的表現，例如：作品等第量表排列法，其以作品量表的方法評定等第。它是包括一系列的樣本作品，依不同的品質高低而給不同的分數，評分者拿它作為參照標準，看作品最接近那個作品水準，就評予那個分數。評分快速簡單，但常流於主觀！

(二)影響實作評量信度與效度的三種評分者誤差來源有：

- 1.教師以編製學習單的方式進行實作評量，但是，在學生完成無數張學習單後，教師評分時可能仍不知道學生是否具備哪些重要的能力，以致評分結果產生偏誤，進而影響信效度。其主因為缺乏具體客觀的評分規準；
- 2.實作評量的目的正是透過模擬真實情境的測量，達到直接測量複雜能力的目的。但是真實目標能力常過於廣泛且複雜，評分者的認知在即是事先設定評分規準下，不免仍有出軌評分的演出，因而降低評量的信效度；
- 3.實作評量的信效度也可能受月暈效應或邏輯錯誤指標之使用的影響。

## 【高分閱讀】

- 1.傅立葉，《教育測驗與統計》總複習班補充閱讀資料一，頁1。
- 2.傅立葉，《教育測驗與統計》高普考考猜五，解釋名詞1、評分規準，頁5-4-5。

二、 $X$ 和 $Y$ 為兩變數，若 $M=aX+b$ 、 $N=cY+d$ 且 $a>0$ 、 $c>0$ 。若 $Q_X$ 、 $S_X^2$ 、 $S_X$ 、 $\bar{X}$ 、 $MDN_X$ 、 $Z_X$ 分別代表 $X$ 的四分差、變異數、標準差、平均數、中位數和標準分數；且 $Q_M$ 、 $S_M^2$ 、 $S_M$ 、 $\bar{M}$ 、 $MDN_M$ 、 $Z_M$ 分別代表 $M$ 的四分差、變異數、標準差、平均數、中位數和標準分數。令 $r_{XY}$ 和 $COV(X, Y)$ 分別代表 $X$ 和 $Y$ 的相關係數和共變數， $r_{MN}$ 和 $COV(M, N)$ 分別代表 $M$ 和 $N$ 的相關係數和共變數。（每一格1分，共25分）

(一)請依序號作答：

$$Q_M = aQ_X + 0$$

$$S_M^2 = (1)S_X^2 + (2)$$

$$S_M = (3)S_X + (4)$$

$$\bar{M} = (5)\bar{X} + (6)$$

$$MDN_M = (7)MDN_X + (8)$$

$$Z_M = (9)Z_X + (10)$$

$$r_{MN} = (11)r_{XY} + 0$$

$$COV(M, N) = (12)COV(X, Y) + (13)$$

(二)若設 $a<0$ 、 $c>0$ 則：

$$S_M^2 = (14)S_X^2 + (15)$$

$$S_M = (16)S_X + (17)$$

$$\bar{M} = (18)\bar{X} + (19)$$

$$Z_M = (20)Z_X + (21)$$

$$r_{MN} = (22)r_{XY} + (23)$$

$$COV(M, N) = (24)COV(X, Y) + (25)$$

答：

(一) (1)  $a^2$

(2) 0

(3)  $a$

(4) 0

(5)  $a$

(6)  $b$

(7)  $a$

(8)  $b$

(9) 1

(10) 0

(11) 1

(12)  $ac$

(13) 0

(二) (14)  $a^2$

(15) 0

(16)  $a$

(17) 0

(18)  $a$

(19)  $b$

(20) -1

(21) 0

(22) -1

(23) 0

(24)  $-ac$

高點 · 高上高普特考

(25) 0

## 【高分閱讀】

傅立葉，《教育測驗與統計》第一回，Ch. 3~Ch. 6，頁10~31。

## 三、請回答下列問題：

(一)請根據你對效度概念的理解，修正以下的說法，使其更正確，請適度說明修正之理由。(10分)

1. 某校之國語成就測驗與某一標準化國語成就測驗成績之相關為0.85，顯示該國語成就測驗是有效度的。
2. 測驗的目標不同，效度因而可以分為三種類別：內容效度、效標關聯效度和建構效度。

(二)某校想知道入學考試的預測效度，因此，計算入學考試成績與入學學生一年後的成績表現得到兩者的相關為0.30，家長們因而抗議該校使用一個效度不佳的測驗篩選學生是不公平的，如果從其他效度證據顯示使用該測驗是適當的，你如何解釋預測效度係數不高的現象？(5分)

(三)在蒐集效度證據上，蒐集測驗得分與其他測量間的相關是常用的方法，請說明在驗證建構效度與效標關聯效度時，測驗得分與其他測量間的相關所扮演的角色有何不同？(10分)

## 答：

- (一)1. 某校的國語成就測驗是依因應測驗需要而新編的工具，某一標準化國語成就測驗理應視為外在效標，兩測驗成績0.85的高相關係數，可以更精確的解釋為該校國語成就測驗具有高的效標(關連)效度，但是，前提要件必須該標準化測驗的外在效標為高品質且適切。
2. 此一說法應修正為：依測驗的種類(內容)不同，為求測驗準確性的高品質，追求的效度有內容、效標、與建構效度三種。例如：成就測驗的重內容效度；性向測驗的重效標效度；以及人格測驗的重建構效度。
- (二)測驗預測效度係數不高的原因，主要是測驗表現與在校成績表現之間的一整年時間間隔，學生的學習成就可能受環境因素或是個人學習態度的影響；也可能在校成績表現來自於品質不佳的測驗，當然無法在兩個成績之間取得高相關係數！
- (三)在驗證建構效度時，測驗得分與其他測量間的相關，又可依性質相同與否，分別檢視與性質不同測量之間低度相關的區別效度，以及與性質相同測量之間高度相關的收斂效度。兩者同時成立才有建構效度的具體證據。而驗證效標關連效度時，只檢視自己使用或編製測驗與性質相同的外在效標之高度相關是否成立。

## 【高分閱讀】

傅立葉，《教育測驗與統計》第三回，第十六章，頁18~19。

## 四、下表為40位大學生男生在身高與體重上的數據。請回答下列問題：

	身高 (cm)	體重 (kg)
平均數	171	65
標準差	4.0	7.0
共變數	11.2	

- (一)身高與體重的相關是多少？如果研究者將身高改為吋(1吋=2.54公分)，將體重改為磅(1公斤=2.2磅)，則身高與體重的變異數、兩者的共變數及相關係數各是多少？(5分)
- (二)請說明下列情形之相關係數是否會改變？並針對影響相關係數的因素提出結論。(5分)
1. 女性樣本之身高與體重的相關為0.30，將男、女樣本合併。
  2. 增加一位身高超過200公分，體重65公斤的男性受試者。
- (三)計算標準迴歸方程式，並解釋方程式中係數的意義。若甲身高的PR值為84，你會預測他的體重是多少公斤？(5分)
- (四)承上題，預測的誤差有多大？如果乙的身高165公分，請根據誤差建立95%的信心區間，並解釋結果。(10分)

答：

$$(一) r = \frac{C_{XY}}{s_X s_Y} = \frac{11.2}{4(7)} = 0.4 \quad \text{該40位大學男生身高與體重的相關為0.4。}$$

身高單位改為吋後，原來以公分計之4.0的標準差，換算為吋之後，取平方得變異數為

$$\left(\frac{4}{2.54}\right)^2 = 2.48；\text{同理，體重改為以磅為單位後，變異數成爲}\left(\frac{7}{2.2}\right)^2 = 10.12。 \text{但是，兩者的共變數與相關係數不變，分別爲11.2與0.4。}$$

(二)1.將男女樣本合併，由於兩性身高與體重的相關因體態與生理的不同，合併後，整合的相關係數一定隨之變動。主因為兩性在身高與體重之標準差的差異。

2.當增加一位超過200公分身高的男性受試者，由於其身高明顯較平均身高高出一大截，然而，體重卻與大多數男同學差不多，計算所得的相關係數應會變小。主要是身高的極端值所造成的影響。

(三)身高與體重的標準回歸方程式為

$$Z_{\hat{Y}} = 0.4Z_X$$

算式中的迴歸係數，正是兩原始變數的相關係數。其表示身高(X)的常態標準化Z值，每變動一個單位，平均而言，會造成對應體重(Y)的常態標準化Z值變動量為0.4。

若某甲的身高PR值為84，則依下列計算，可推斷其體重為：

$$p(X \leq x) = 0.84$$

$$p(Z \leq 1) = 0.84$$

$$\frac{x-171}{4} = 1 = Z_X$$

$$Z_{\hat{Y}} = 0.4Z_X = 0.4(1) = 0.4$$

$$\text{因此，} 0.4 = \frac{y-65}{7} \quad y = 67.8(kg)$$

(四)預測的誤差，正是估計標準誤的概念。其值可以計算如下式：

$$s_{Y.X} = s_Y \sqrt{1-r^2} = 7\sqrt{1-0.4^2} = 6.416$$

若某乙的身高為165公分，則其體重的預測值為

$$Z_X = \frac{165-171}{4} = -1.5$$

$$Z_{\hat{Y}} = 0.4(-1.5) = -0.6$$

$$\frac{y-65}{7} = -0.6 \quad y = 60.8(kg)$$

$$60.8 \pm t(0.025, 38) \left[ 1 + \frac{1}{40} + \frac{(165-171)^2}{4^2(40-1)} \right] (6.416)$$

$$\begin{aligned} \text{其體重95\%的信賴區間爲} &= 60.8 \pm 2.021(6.947) \\ &= 60.8 \pm 14.04 \\ &= (46.76, 74.84) \end{aligned}$$

此一區間表示，在100次抽樣中，約有95次可以利用乙身高的數據，以及該回歸模式，正確的將乙的體重以該區間估計。

#### 【高分閱讀】

1. 傅立葉，《教育測驗與統計》總複習第一回，頁78，例題6；頁80，例題12；解答公式，頁110與頁113。
2. 傅立葉，《教育測驗與統計》第一回，頁76，例題12。