

《教育測驗與統計》

- 一、某校在數學(M)與英語(E)的評量方式包括教室觀察(O)、紙筆測驗(T)以及實作表現(P)等三種，分別以MO、MT、MP、EO、ET、EP代表學生在這三種評量方式中，數學和英語的得分情形。該校計算學生這6項分數之間的兩兩相關。請問若要證明這些評量有聚斂(Convergent)及區辨(Divergent)效度，則這些分數中，那些相關應該較高？而那些相關應該較低？
- (一)請先說明何謂聚斂(Convergent)及區辨(Divergent)效度。(15分)
- (二)再說明那些應較高，那些應較低。(10分)

試題評析	本題考試主題為教育測驗的效度。而且是以構念效度在多項特質-多項方法矩陣(MTMM)所離析出來的聚斂與區辨兩個效度為主。題目簡單易答，得分關鍵在於第二小題的陳述展開方式。
考點命中	1.《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，第十六章，頁27。 2.《高點·高上教育測驗與統計總複習講義》，傅立葉編撰，頁6，三、最新趨勢。 3.《高點·高上教育測驗與統計重點題神》，傅立葉編撰，第八題。

答：

- (一)本題針對數學與英語兩個科目能力的評量分別採取教室觀察、紙筆測驗與實作表現等三種方式。對於評量方式或工具之品質的探討，特別是在效度上，常用聚斂效度與區辨效度的概念加以進行。兩者的意義分別說明如下：
- 1.聚斂效度：當測量相同構念的不同工具間，測量結果之表現或分數存在高度相關，則稱其中之新編測驗或目標評量工具具有聚斂效度。或是，當測量同一構念之多重指標彼此間存在高度關聯時，則此效度存在。甚且，評量工具欲測量之特質與理論上相關特質的關聯程度也稱為聚斂效度。
 - 2.區辨效度：評估工具能否區辨不同能力之受測者間差異的效度。也可以說是當所欲測量之構念的多重指標應與對立之構念的測量指標之間存在低度或甚至負向相關的效度概念。
- 當評斷一個新編測驗或評量工具之效度品質時，如果藉由多項特質-多項方法矩陣(MTMM)進行分析，能夠同時得出聚斂與區辨效度的證據，則該測驗或評量工具具有高的構念效度，評量結果對於受試者的能力與特質較為準確。
- (二)1.當分別採取教室觀察(O)、紙筆測驗(T)與實作表現(P)等三種方式進行數學(M)與英語(E)兩個科目能力的評量，然後計算兩兩之相關，利用多項特質-多項方法矩陣可以將上述之兩兩相關係數呈現如下：

	教室觀察(O)		紙筆測驗(T)		實作表現(P)	
	數學(M)	英語(E)	數學(M)	英語(E)	數學(M)	英語(E)
<u>教室觀察(O)</u>						
數學(M)	MM	ME	MM	ME	MM	ME
英語(E)	ME	EE	ME	EE	ME	EE
<u>紙筆測驗(T)</u>						
數學(M)	MM	ME	MM	ME	MM	ME
英語(E)	ME	EE	ME	EE	ME	EE
<u>實作表現(P)</u>						
數學(M)	MM	ME	MM	ME	MM	ME
英語(E)	ME	EE	ME	EE	ME	EE

- 2.當以MO、MT、MP、EO、ET與EP分別表示學生在三種評量方式中的兩個科目能力的得分，由於數學與英語兩個科目能力的評量著重之表現方式不同，理論上數學應重紙筆測驗與實作表現，而英語應重教室觀察與紙筆測驗，因此，MT與MP、EO與ET的分數相關應該較高，而MO應與MT及MP的相關較低，且EP的分數應與EO及ET的相關較低。

【版權所有，重製必究！】

二、請問何謂效標參照測驗？何謂常模參照測驗？請問大學學科能力測驗與補救教學科技化評量是屬於那種測驗？請說明你的理由。(20分)

試題評析	本題涉及教育測驗對於測驗的一種大分類的兩種測驗之定義與舉例判斷。不是意外的困難考題，考生大致應能順利應對，取得本題的大多數分。
考點命中	1.《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，第十四章，頁6-7。 2.《高點·高上教育測驗與統計總複習講義》，傅立葉編撰，頁4，三、最新趨勢；頁83-84，三、最新考試重點補充。

答：

(一)效標參照測驗與常模參照測驗是測驗學的一種分類，茲分別說明如下：

1.效標參照測驗：

強調個體與某一效標(標準)的比較，著重學生學習結果是否達成教學目標。由於行為目標的發展，教學目標可以明確的條列，教學績效端視兩者相符的程度而定。又稱標準參照評量，因事前已決定所要達到之標準或學習結果，其分數不因參照團體的改變而不同，每一教學活動亦需達到一定的標準才能停止，教師教學的責任也才算是盡到，是故較易維持一定的水準。效標具備事前主觀決定之絕對標準的特性。常可與非標準化測驗或非正式測驗為同義詞。如一般教師自編測驗與隨堂測驗。

2.常模參照測驗：

指測驗結果根據個別受試者分數在團體中的相對表現位置加以解釋的一種測驗。比較或參照的標準為事後相對的平均水準，客觀評斷表現好壞的能力較強。又分為：團體內常模與發展性常模。常可與標準化測驗或非正式測驗為同義詞。如大型升學或國家考試，以及智力或性向測驗。

(二)大學學科能力測驗因為屬於升學考試之一，個別考生的能力水準依全體考生在該項測驗的平均水準進行相對表現好壞的評斷，使用的是標準分數常模或百分等級常模的團體內常模，因而屬於常模參照測驗。至於，補救教學科技化評量多屬學習過程中，教學者依需要所實施之評量，以取得學生學習進步或解決問題能力提升的訊息，其所判斷的標準多為教學者事先主觀設定，應屬效標參照測驗。

三、有一個研究者其研究所得平均值為103， $N=36$ ，已知 $\sigma=12$

當 $H_0: \mu=96$ (附常態分配表)

(一)當其設定 $\alpha=.05$ ，統計考驗力為何？(7分)

(二)當 N 增加至64人，在 $\alpha=.05$ 的情形下，統計考驗力為何？(8分)

試題評析	本題涉及推論統計之統計考驗力的定義、計算與常態機率查表。只要清楚型II錯誤的定義與常態分配表的查表邏輯，並利用信賴區間估計的機率式，應能順利解答！正因為統計考驗力的計算題相對於傳統申論問答的命題冷門，因此只有15分的配分。
考點命中	《高點·高上教育測驗與統計講義》第二回，傅立葉編撰，第十四章，頁6。

答：

依題意，該研究者所得到的統計資料，在 $\mu=96$ 的虛無假設下，分別計算並回答以下子題：

(一)當 $\alpha=0.05$ ，基於 $N=36$ 與 $\sigma=12$ 之大樣本且母體標準差的條件，在此對於母體平均數的假設檢定應採取常態分配 z 檢定。惟，考驗力分析可以先利用區間估計方式，定義犯型II錯誤(β)的機率式如下：

$$\begin{aligned}\beta &= p(96 - 1.96(12)/\sqrt{36} \leq \bar{x} \leq 96 + 1.96(12)/\sqrt{36} | \mu \neq 96) \\ &= p(92.08 \leq \bar{x} \leq 99.92 | \mu \neq 96)\end{aligned}$$

接著，利用假設之真實值102可以進一步計算統計考驗力如下：

$$\begin{aligned}\beta &= p(92.08 \leq \bar{x} \leq 99.92 | \mu = 102) \\ &= p\left(\frac{92.08 - 102}{12/\sqrt{36}} \leq z \leq \frac{99.92 - 102}{12/\sqrt{36}}\right)\end{aligned}$$

【版權所有，重製必究！】

$$\begin{aligned}
 &= p(-4.96 \leq z \leq -1.04) \\
 &= 0.1492 - 0 \\
 &= 0.1492
 \end{aligned}$$

因此檢定考驗力為 $1 - \beta = 0.8508$

(二)其他條件不變，當N增加為64人，同前題可以依序計算與查表如下：

$$z = \frac{103 - 96}{12/\sqrt{64}} = 4.666 \cong 4.67$$

$$\begin{aligned}
 \beta &= p(96 - 1.96(12)/\sqrt{64} \leq \bar{x} \leq 96 + 1.96(12)/\sqrt{64} | \mu \neq 96) \\
 &= p(93.06 \leq \bar{x} \leq 98.94 | \mu \neq 96)
 \end{aligned}$$

接著，利用假設之真實值102可以進一步計算統計考驗力如下：

$$\begin{aligned}
 \beta &= p(93.06 \leq \bar{x} \leq 98.94 | \mu = 102) \\
 &= p\left(\frac{93.06 - 102}{12/\sqrt{64}} \leq z \leq \frac{98.94 - 102}{12/\sqrt{64}}\right) \\
 &= p(-5.96 \leq z \leq -2.04) \\
 &= 0.0207 - 0 \\
 &= 0.0207
 \end{aligned}$$

因此檢定考驗力為 $1 - \beta = 0.9793$

四、某研究者想利用路徑分析探究自我效能與學習成就的關係是否受到學習態度中介變項的影響。其路徑假設如下圖，而研究者共做了4個迴歸分析如下表所示：

(一)請問圖中A、B、C要填上那些數字？(15分)

(二)根據路徑分析的結果，你認為自我效能與學習成就的關係是否透過學習態度中介變項的影響？請說明理由。(10分)

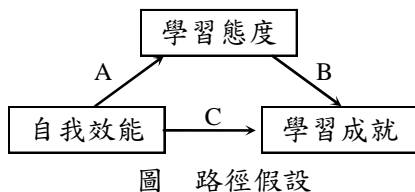


表 迴歸分析資料表 (N=107)

依變項	自變項	Beta	R ²
學習態度	自我效能	.758**	.575**
學習成就	自我效能	.309**	.095**
學習成就	學習態度	.393**	.154**
學習成就	自我效能	.373**	.155**
	學習態度	.027	

試題評析	本題涉及中介變項在因果關係中的角色與迴歸分析，是以往教育心理研究所入學考試研究法的教育研究法範圍。平心而論，有點超過命題範疇，也是今年整份測驗最難的題目，但是，如能運用迴歸分析的基本概念，在第一小題將簡單迴歸分析之標準化迴歸係數視為路徑係數進行回答，仍能取得可觀的分數，對於高分錄取具有關鍵性影響。
考點命中	《高點·高上教育測驗與統計講義》第一回，傅立葉編撰，第八章，頁53。

答：

依題意，路徑分析應用於自我效能與學習成就的關係是否受到學習態度之影響的迴歸分析中，以下分別回答各子題：

(一)路徑假設圖中之A、B、C分別可填入0.758、0.393與0.309等三個標準化簡單迴歸分析模式之係數Beta值。

(二)中介變項的影響效應一般可以分為全部與部分效應兩種。依題目中所進行迴歸分析結果的數據，因為自我

效能對於學習成就的C路徑係數達顯著，但是，在第四個迴歸分析中之學習態度對於學習成就的路徑係數0.027未達顯著，因此，自我效能與學習成就的關係並未受到學習態度之影響，也就是說，學習態度之中介變項的假設未能影響自我效能與學習成就的因果關係，此迴歸分析中之學習成就直接受到自我效能的顯著影響。

五、某研究者從社會階層低、中、高分配依序是25%、60%、15%的區域隨機抽樣200個樣本進行電話調查訪問。其樣本社會階層的分布如下表：

	低	中	高
人數	37	130	33

請問這個樣本的社會階層分布是否顯著地異於母群體？若統計檢定的臨界值設為A，請進行統計檢定並下結論。（15分）

試題評析	本題涉及類別變項資料的卡方檢定法。只要根據題意在一開始即釐清屬於適合度考驗，後續簡單的計算與檢定步驟及結論則無太大挑戰。
考點命中	《高點·高上教育測驗與統計講義》第二回，傅立葉編撰，第十二章，頁35-37。

答：

依題意，低、中、高的社會階層分配因為屬於類別變項(名目尺度)，對於樣本之社會階層分布是否異於母群體之研究問題的假設檢定應該採用卡方統計法之適合度檢定。

(一)首先，虛無假設與對立假設為：

$$H_0: p_1 = 0.25 \quad p_2 = 0.60 \quad p_3 = 0.15$$

(二)將理論人數與觀察人數整理，得下表：

	低	中	高
觀察人數	37	130	33
理論人數	50	120	30

(三)計算檢定統計量

$$\chi^2 = \frac{(37-50)^2}{50} + \frac{(130-120)^2}{120} + \frac{(33-30)^2}{30} = 3.38 + 0.833 + 0.3 = 4.513$$

(四)臨界值可以利用最常用之0.05的顯著水準與常態分配表而得

$$\chi^2_{2} = 5.991$$

(五)檢定結論

因為4.513的統計量小於5.991的臨界值，檢定結論為不拒絕虛無假設，意即在0.05的顯著水準下，利用這200個隨機樣本的資訊無法提供充分證據足以結論樣本的社會階層分布顯著異於母體。

【版權所有，重製必究！】