

《教育測驗與統計》

試題評析	今年考題難度偏低，特別是教育統計學部分，艱難的數學公式與計算並未出現，應該可以讓大多數考生感到寬心。測驗學的信度、難度、鑑別度以及誘答力都是重點中的精華，考生大多有準備；統計學在第三題考的檢定步驟其實沒有絕對的順序，只要前後步驟大致合理，說法合乎邏輯，應不致會有嚴重扣分；第四題的一因子變異數分析電腦報表解讀也相當容易。但是，正因為難度不高，考生必須要獲得相對高分才能在最後脫穎而出。整體而言，程度好的考生應能取得85分以上的高分。
------	--

一、測驗是研究者用來收集受試者作答反應資料的工具，請問：

- (一)何謂測驗的信度(reliability)? (5分)
- (二)常模參照測驗(norm-referenced test)的信度係數可分成那幾種? (8分)
- (三)承(二)，每一種信度係數的主要誤差來源各為何? (8分)
- (四)常在公務人員考試中使用的測驗式試題，如果考試後想進行試題分析的話，它最適合使用那一種信度係數作為分析指標? (4分)

試題評析	本題以測驗學的信度為命題重心，並以常模參照測驗作為各種可能信度的問答與要求回答誤差來源，有準備的考生答題時應當不太有負擔與阻礙。
考點命中	1.高上《教育測驗與統計講義》第三回，傅立葉編撰，第十四章，頁11-14。 2.高上《102地特考猜-教育測驗與統計》，傅立葉編撰，頁1。

答：

- (一)測驗的信度指測驗分數所提供受試者資料的可靠性，也代表一測驗在測量某心理特質或某學科學習成就上所呈現的一致性或穩定性。因測驗信度的高低，測驗結果給予受試者對分數的信心大小也就不同。測驗的信度可以由組成試題在測量特質或能力之目標的相符或一致程度加以探討；或者，透過兩次前後測驗分數高低的一致性進行解讀。以數量方法來說，則可藉由相關係數或是測量標準誤的計算進行掌握。
- (二)常模參照測驗通常指的是標準化測驗，也是正式測驗的同義詞。探討這類測驗的信度係數，常可分為：
 - 1.重測信度，又稱再測信度，使用同一測驗，針對同一群受試者，在不同時間內前後測兩次，再求兩次得分間的相關係數(可依測驗結果分數資料類型，選擇計算積差相關係數或等級相關係數)，即得。
 - 2.複本信度，也稱複份信度，針對目標特質或能力編有就內容、性質、型式、題數和難度等皆大致相近的兩份試卷，利用這兩份試卷所得分數計算所得的相關係數，即稱。又可因兩份測驗的施測時間是否存在區隔，可再分為等值係數與穩定且等值係數兩種信度概念。
 - 3.折半信度，又稱內部一致性信度，在施測一份測驗後，將測驗題目依奇偶折半或隨機折半等適當的方式分為兩半計分，再求兩半分數之相關係數而得。由於題數減半，通常求得的信度係數傾向於低估，還要再利用校正公式加以還原，以估計原本整份測驗的真正信度。
- (三)呈(二)，三種信度的誤差來源分別為：
 - 1.重測信度：時間取樣(指的是測驗施測間隔時間與否以及長短的選擇與操作)
 - 2.等值信度：內容取樣(指的是複本試題組合的同質性與一致性)
 - 3.穩定且等值信度：時間與內容取樣
 - 4.折半信度：內容取樣(受折半方式的不同，計算信度的兩半測驗組合不同而影響信度估算)
- (四)常在公務人員考試中使用的測驗式試題，應以選擇題為主要題型，因為測驗題數動輒30-50題，若考後想進行試題分析，最適合做為分析指標的信度係數應為Cronbach alpha信度，其計算多數試題得分間的複相關係數。

二、某一位研究員在針對某次測驗後進行試題分析時，獲得下表有關某一試題的各選項選答資料。

組別	選項				
	*A	B	C	D	未答
高分組	20	10	50	20	0
低分組	10	25	25	40	0

註：細格中的數字表示「人數」，*表示「正確答案的選項」。

請參考上表，逐一回答下列問題：

- (一)請計算該試題的難度指標值？(4分)
- (二)請計算該試題的鑑別度指標值？(4分)
- (三)請評估該試題是否具有理想的選項誘答力？(5分)
- (四)請評鑑該試題的優劣，並說明你的結論。(6分)
- (五)假如該名研究員發現該試題的答案可能是錯誤的。基於該測驗中每一試題都具有優良試題特質的前提下，請問該試題的正確答案應該是那一個選項才較為合理？請說明你的理由。(6分)

試題評析	本題命題以測驗學的「試題分析」為標的，答題核心在於難度與鑑別度指標值的計算，以及兩者數據大小關係在判斷試題品質的說明。難度不高，多數考生應能夠高分掌握全數配分。
考點命中	1.高上《教育測驗與統計講義》第三回，傅立葉編撰，第十八章，頁32-33；觀念練習題Q20，頁37-38，模擬試題，Q11，頁40。 2.高上《102地特考猜-教育測驗與統計》，傅立葉編撰，頁14-15，5.難度與鑑別度分析。

答：

(一)該試題的難度指標值可以計算如下：

$$P = \frac{\frac{20}{50} + \frac{10}{50}}{2} = 0.3$$

(二)該試題的鑑別度指標值可以計算如下：

$$D = \frac{20}{50} - \frac{10}{50} = 0.2$$

- (三)以試題的誘答力而言，應以低分組在每一個錯誤選項的選答人數多於高分組，以及每一個錯誤選項至少可以吸引一位低分組受試者選答為判斷依據。因此，以正確答案A以外的B、C、D三個錯誤選項來看，只有C選項高分組比低分組答題人數多而較不理想外，錯誤選項B與D的低分組選答人數皆高於高分組，因此，整體而言，該試題具有不錯的誘答力。
- (四)試題的優劣可由其難度與鑑別度指標值加以判斷。依(一)(二)計算而得的兩個指標值的數據，0.3所表示高低分組平均答對比例所指的難度，指出該題難度偏高；因而，導致0.2的鑑別度也較差，所以就兩大指標而言，該試題的品質不是太好。
- (五)當A的正確選項答案是錯的，以及在優良試題的假設前提下，正確答案應以選項C較為合理，因為高分組在正確選項選答的人數應高於低分組人數，才能符合試題得分與總分間之高度一致性的優良試題條件。

三、在教育統計學中，某研究員若想檢定「數學考試焦慮是否存在著性別差異」的假設，他該怎麼做，才比較恰當？請說明你的檢定步驟。(25分)

試題評析	本題考的是推論統計的基礎概念，題目並未提供任何數量，答題重點在於確認是兩個獨立母體平均數的差異顯著性檢定，並勾勒出可選擇t-分配或Z-分配。然後依序回答檢定之步驟。題目難度不高。
考點命中	1.高上《教育測驗與統計講義》第二回，傅立葉編撰，第十一章，頁19-21。

2.高上《102地特考猜-教育測驗與統計》，傅立葉編撰，頁6表格。

答：

當研究員想要檢定數學考試焦慮(程度)是否存在性別差異，因為涉及男性與女性兩個母體，而考試焦慮若以量表取得數據，最適當探討的統計量為平均數，因此，在兩個母體平均數差異顯著性的研究問題與檢定工作中，

除了必須先定義研究母體、抽樣兩性學生進行數學考試焦慮測驗，並決定顯著水準後，檢定步驟如下：

- 1.寫出虛無假設與對立假設(此題為雙尾檢定)
- 2.利用顯著水準，找出臨界值
- 3.決定決策法則(拒絕虛無假設時，檢定統計量與臨界值的大小關係)
- 4.計算兩性的樣本平均數與相關統計量
- 5.計算檢定統計量(通常就t-分配與常態分配進行選擇)
- 6.進行檢定結論
- 7.探討犯型I與型II錯誤的可能與機率

四、參考下列某個實驗結果的電腦報表數據：

SV	SS	df	MS	F	p
Between group	1200	2	600	300.0	.000
Within group	50	25	2		
Total	1250	27			

請回答下列問題：

- (一)總共有幾名受試者參與實驗？(2分)
- (二)這些受試者共分成幾組進行實驗？(3分)
- (三)本實驗的虛無假設應該如何表示？(5分)
- (四)本實驗的統計檢定結果為何？(5分)
- (五)請你解釋這個實驗的結論。(10分)

試題評析	本題命題之統計方法為一因子變異數分析，以提供電腦報表之完整數據，考生須具備報表讀表能力。除能利用自由度的簡單計算回答(一)(二)兩題外，也須了解p-值檢定法的決策法則，才能掌握(四)(五)兩小題的分數。本班考生對於此題的回答毫無困難。
考點命中	1.高上《教育測驗與統計講義》第二回，傅立葉編撰，第十三章，頁42-43。 2.高上《102地特考猜-教育測驗與統計》，傅立葉編撰，頁7-8，變異數分析摘要表。

答：

(一)依題目提供的電腦報表數據，可以利用自由度(df)欄位中之Total的27計算得知：

$$27 = n - 1$$

$$n = 28$$

因此，參與本項實驗的受試者共有28名。【版權所有，重製必究！】

(二)同理，利用自由度(df)欄位中之Between Group的2計算得知：

$$2 = k - 1$$

$$k = 3$$

因此，這些受試者共分成三組進行實驗

(三)本實驗的虛無假設應表示為

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

(四)本實驗的統計檢定結果，依電腦報表數據可得：

因.000的p-值小於任何常用的顯著水準，所以檢定結果為拒絕虛無假設，結論對立假設

(五) 依前一小題的檢定結論，可以解釋為：

利用這28名受試者的樣本資料，在隨機分派下，三組受試者的平均數不全等，也就是檢定達顯著。意即，各組平均數之間可能存在兩兩差異，如果實驗目的在於了解三種不同教學方法，就某一特定學科，教學與學習成效的差異，結論是確實提供顯著證據足以推論成效並不全然相同，其也表示某一特定實驗處理(分組)的學習效果要顯著高於其他分組。

高點 · 高上

【版權所有，重製必究！】