

《教育測驗與統計概要》

試題評析

今年的考題偏易，因為平日上課的講解與例題的練習皆有出現，本班學生在第二、三、四題的作答，應能全部掌握。估計程度好、準備周到、臨場發揮實力的考生應可輕易取得80分以上。

第一題：性向測驗(教材Ch.14, p.2)；真實性評量(即Ch.19, p.37的實作評量)

第二題：在本班教材Ch.18, p.24)完全命中。

第三題：在教材Ch.4, p.13與Ch.5, p.18可找到計算公式。本題要拿滿分的關鍵在於，第三小題考生能否依題意，將CV值依大小排列並加以解釋其意義。

第四題：在教材的Ch.7, p.28-29與Ch.8, p.40完全命中！最困難的應是第二小題，如能回答詳細，條理有秩，應能取得高分！

一、請解釋下列有關測驗的名詞：（每小題5分，共25分）

(一)多元智能(multiple intelligence)

(二)性向測驗(aptitude tests)

(三)真實性評量(authentic assessment)

(四)課程本位能力測驗(curriculum-based competency tests)

(五)工作取樣系統(work sampling system)

答：

(一)多元智能(multiple intelligence)

智能是一種人性整合的生活操作模式，也是解決問題或創造的能力，而不能只側重IQ。但智能並不是與生俱來的，每個人都有能力改進且擴展自己智能。每個人的智能是多元的，並有自己獨特的智能組合。根據哈佛大學心理學家迦納博士(Howard Gardner)的理論，每個人都具備多元智能的潛力，包括：語言、邏輯-數學、空間、肢體-運作、音樂、人際、內省與自然等智能。

(二)性向測驗(aptitude tests)

性向測驗為發現受試者個人的能力傾向與長處，以作為其指導升學或選擇職業的依據，並且協助施測者達到因材施教或人盡其才的職業輔導為目的的測驗。它又可分為用以鑑別個人多方面特殊潛在能力的綜合性向測驗，其用多個測驗分數的結果來推估個人在某方面性向的高低；以及用以鑑別個人具有如文書、機械性向等特殊潛在能力的特殊性向測驗等兩大類。

(三)真實性評量(authentic assessment)

當直接去測量學生在某一課程的實際操作的表現，以評量學習者的教育成就的評量，就稱之為真實性評量。又有人稱之為實作評量(performance assessment)、或直接評量(direct assessment)。其評量方法可以包括：書寫測驗、問題解決、實驗操作、展示、表演、作品集、教師觀察、檢核表、問卷，以及團體合作計畫等技術。而其實施方式可以是日常教室的形成性評量、總結性評量或特殊教學計畫，其評鑑是以人的專業判斷為主，並強調多元化評量。真實性評量所要評量的是學生所欲達成的學習目標，以及直接教學的結果。它適用於各年齡層、各學習階段的各學科學習評量。對於教師，其可以改進與延伸教學目標；對於學生，則可以評估其學習的進步情形。

(四)課程本位能力測驗(curriculum-based competency tests)

每一個科目或課程，都有其教材大綱及教學重點，同時也要考慮教學目標。因此，在評量學生的課程本位能力時，應針對該科學習的活動，考慮包括教材內容、行為目標、學習情境、以及學生個人的背景能力等各方面，即同時考慮課程設計、課程實施、學習情境、個人等因素。

(五)工作取樣系統(work sampling system)

即從工作項目中找出重要的樣本，讓學生操作反應，再進行觀察和記錄的一種有系統編製測驗的方式。

【參考書目】

1. 性向測驗(教材Ch. 14, p. 2)；真實性評量(即Ch. 19, p. 37的實作評量)

二、在以古典測驗理論(classic testing theory)進行試題分析(item analysis)時，通常要計算每一個個別試題之難度(item difficulty)與鑑別度(item discrimination)，並檢視試題選項(options)的分布，也就是誘答項分析(distractor/distracter analysis)。請回答下列有關試題分析的問題：

- (一)請分別說明試題之難度與鑑別度的意義。(5分)
- (二)請說明試題之難度與鑑別度的關係。(10分)
- (三)何謂誘答項分析？請以四個選項的單選選擇題為例說明之。(10分)

答：

- (一)試題的難度，最簡單的意義是指所有受試者，在個別試題的答對人數比例；當依總分區別高低分組時，則難度指的是高低分組在特定試題答對人數比例的平均數。至於鑑別度指的是個別試題鑑別受試者的程度之能力高低，其可以用高低分組在特定試題答對人數比例的差異求得。
- (二)通常個別試題的難度趨於適中(即 $P=0.5$)，則其鑑別度愈高；當試題的難度過高或過低，則其鑑別度通常會較低。
- (三)以四個選項的單選選擇題為例，誘答項分析指的是在正確答案(選項)以外，其餘三個錯誤答案引誘受試者圈選的能力高低之探討。誘答力夠高的試題選項，常是用以評鑑試題品質的指標之一。至於如何判斷選項具有誘答力？通常有兩個條件：
 1. 至少一位低分組學生選擇任何一個錯誤選項；
 2. 選擇錯誤選項的低分組學生較高分組學生多。

【參考書目】

1. 本班教材(Ch. 18, p. 24)完全命中。

三、某生在第一學期結束後，其所修的五個科目的成績及其全班在此五個科目成績的平均數及標準差分別如下：

科目	某生成績	全班平均數	標準差
國文	85	83	7
英文	82	81	9
數學	75	70	10
物理	73	71	11
化學	80	78	8

- (一)請根據上表分別計算該生五科成績之標準分數，也即 z 分數。(10分)
- (二)若該生與全班同學比較，其最強與最弱的科目是那兩個科目？(5分)
- (三)平均數及標準差的應用之一是用於計算變異係數(coefficient of variation；簡稱CV)，也即標準差相對於平均數的比，並以百分比表示；因此，CV可用來比較單位不同的變項之相對變異情形。請依以上描述，計算上表五個科目的相對變異情形，並由高至低分別列出各科目之CV。(10分)

答：

(一)該生五科成績的標準 z 分數分別計算如下：

$$\text{國文} \quad z = \frac{85 - 83}{7} = 0.2857 \cong 0.286$$

$$\text{英文} \quad z = \frac{82 - 81}{9} = 0.1 \cong 0.111$$

$$\text{數學} \quad z = \frac{75 - 70}{10} = 0.5$$

$$\text{物理 } z = \frac{73 - 71}{11} = 0.1818 \cong 0.182$$

$$\text{化學 } z = \frac{80 - 78}{8} = 0.25$$

(二)相對於全班其他同學，該生最強與最弱的科目，可以藉由在子題(一)所計算的相對位置量數 z 值的大小得知，因數學科的 z 值為五科中最大的0.5，以及因英文科的 z 值為五科中最小的0.111，因此數學科為其最強，而英文科為其最弱的科目。

(三)依該班五個科目的平均數與標準差所計算的變異係數分別為：

$$\text{國文 } CV = \frac{7}{83} = 0.0843 = 8.43\%$$

$$\text{英文 } CV = \frac{9}{81} = 0.1111 = 11.11\%$$

$$\text{數學 } CV = \frac{10}{70} = 0.1429 = 14.29\%$$

$$\text{物理 } CV = \frac{11}{71} = 0.1549 = 15.49\%$$

$$\text{化學 } CV = \frac{8}{78} = 0.1026 = 10.26\%$$

由高而低列出為：物理(15.49%)、數學(14.29%)、英文(11.11%)、化學(10.26%)、國文(8.43%)，其表示該班學生的成績在物理科變化(異)最大，而在國文科的變異最小。

【參考書目】

1. 本班教材(Ch. 4, p. 13與Ch. 5, p. 18)可找到計算公式。

四、皮爾遜積差相關(Pearson's product-moment correlation, 以 r 表示)係最常用的統計分析方法之一。請回答下列有關皮爾遜積差相關的問題：

- (一)從雙變項資料(bivariate data)的關係及測量量尺(scale of measurement)的觀點，說明使用此種分析方法的條件？(5分)
- (二)解釋積差相關係數(r)時要注意那些事項或限制？(10分)
- (三)若有一項研究結果指出IQ與學業成就的積差相關係數(r)為0.90，從變異的觀點，要如何解釋IQ對成就的貢獻有多少？(5分)
- (四)因為前項研究結果並非完全正相關，從變異的觀點，還有多少變異無法解釋？(5分)

答：

- (一)使用皮爾遜積差相關進行統計分析，雙變項資料的測量量尺應至少為區間(等距)尺度，也適用於比率(等比)尺度，也就是所謂的連續型變項資料。
- (二)因為計算所得的相關係數值，其大小與正負影響對於兩個變項之間相關程度與情形的了解，在解釋其值時應注意的事項有：
 1. 積差相關係數的值域介於-1與1之間，通常愈趨近於正負1的數值，表示兩變項間的關聯程度愈高，差別只是愈趨近正1，表示兩變項間為強烈正相關；而當其值愈趨近負，則表示兩變項間為強烈負相關。
 2. 但是積差相關係數值常易受計算的資料筆數(或為樣本量)多少的影響，在解讀兩個高低大小不等的相關係數值時，應注意其背後所使用的樣本量大小，在相當小的樣本量下，即使相關係數值非常大，研究者也不見得高興的起來；但是當樣本量非常大，則即使相關係數值小至0.4，都可能是對於研究目標的兩個變項之強烈(甚至顯著)相關的重要資訊。

3.即使計算得到表面上相當接近正負1的相關係數值，也不必然具有推論上的實質價值與能力，仍必須透過假設考驗，進行其統計顯著性的再確認！

4.兩變項間強烈的正(負)相關，不一定表示因果關係的成立。仍須由滿足常理、邏輯、理論，以及發生時間的落差性，加上對於可能存在的干擾、調節變項的檢視，才可進行因果推論。

(三)由於IQ與學業成就的相關係數(r)為0.9，可以利用迴歸分析之變異的觀點，以決定係數(R^2)之可解釋變異佔總變異的比例，即 $R^2 = (r)^2 = 0.9^2 = 0.81 = 81\%$ ，獲得IQ對於成就的變異的解釋能力，具有81%的貢獻。

(四)以100%減去子題(三)的81%，可知仍有19%的學業成就之變異無法由IQ與對應的迴歸模式加以解釋。

【參考書目】

1. 本班教材(Ch. 7, p. 28-29與Ch. 8, p. 40)完全命中！

