

《教育測驗與統計》

一、王老師要瞭解啟發教學法、欣賞教學法及建構式教學法，在不同學習壓力（區分為高度及低度壓力）的學習表現之影響情形。每一種教學法各十位學生，每種教學法在高度及低度壓力各五名學生。經過教學實驗之後，透過二因子變異數分析，結果摘要如表。請回答以下問題：

(一) 請寫出 A 至 K 格中的數值（請以 A = ..、B = ... 依此類推，寫於答案紙）。（11 分）

(二) 表中看出那些項目達到統計顯著水準呢？為什麼？因為有些項目達到統計顯著水準，所以王老師下一個步驟應進行那些統計檢定呢？表中有一項為交互作用，請說明它的種類及其特性。（14 分）

表：壓力與教學方法對於學習表現的二因子變異數分析摘要

變異來源	SS (離均差平方和)	df (自由度)	MS (均方)	F 值
組間				
壓力	4.0	A	E	I
教學方法	100.0	B	F	J**
壓力與教學方法交互作用	40.0	C	G	K*
組內 (誤差)	96.0	D	H	
總和	240.0			

* $p < .05$ ** $p < .01$

試題評析	本題的命題環繞在二因子變異數分析，沒有要求考生使用複雜的原始資料進行計算，而是要利用 ANOVA 表的數學關係進行英文字母編號之空格的簡要計算；另外，考生必須使用統計軟體報表常用以表示檢定達顯著的*號回答第二個問題，並針對交互作用達顯著的意義與後續統計分析作更深入的敘述。
考點命中	1.《高點·高上教育測驗與統計講義》第二回，傅立葉編撰，頁51-52、56-57。 2.《高點·高上教育測驗與統計講義》總複習第一回，傅立葉編撰，頁4。

答：

(一) 針對王老師就不同教學方法在不同學習壓力下的學習表現影響情形的研究問題，經教學實驗後進行二因子變異數分析所得之變異數分析摘要表中的 A 至 K 空格中的數值，分別計算回答如下：

A=1 (高低兩種壓力的自由度為 $2-1=1$)

B=2 (三種教學方法的自由度為 $3-1=2$)

C=2 (壓力與教學方法之交互作用的自由度為 $1(2)=2$)

D=24 (組內自由度為 $29-1-2-2=24$)

E=4 (壓力均方為壓力離均差平方和除以自由度, $4/1=4$)

F=50 (教學方法均方為教學方法離均差平方和除以自由度, $100/2=50$)

G=20 (壓力與教學方法交互作用均方為 $40/2=20$)

H=4 (誤差均方為 $96/24=4$)

I=1 (壓力的 F-值為 $4/4=1$)

J=12.5 (教學方法的 F-值為 $50/4=12.5$)

K=5 (交互作用的 F-值為 $20/4=5$)

(二) 1. 表中達到顯著水準的有教學方法的 F 值(J)，因為兩個星號表示在 0.01 的顯著水準下，12.5 的 F 值拒絕虛無假設而達顯著；以及壓力與教學方法交互作用的 F 值(K)，因為一個星號表示在 0.05 的顯著水準下，5 的 F 值拒絕虛無假設而達顯著。

2. 由於壓力與教學方法的交互作用檢定達顯著，顯示壓力對學習表現的影響，會受到教學方法的牽動，而

教學方法對於學習表現的影響，會受到壓力的牽動，此時自變項教學方法主效果達顯著的檢定結論已無法符合我們的需求，進一步的作法應固定壓力之各水準(level)，進行教學方法之one-factor ANOVA；另也需要固定教學方法之各水準(level)，進行壓力之one-factor ANOVA。如此總共需要進行五個額外的假設檢定。

- 3.交互作用是二因子變異數分析方法的一個重要特性，其提供主要自變項個別對於依變項的主效果之外的一個重要資訊，不只讓研究議題在兩個自變項對於依變項的因果關係探究下，提供更有價值的訊息而讓分析或檢定結論更加充實，更能深入瞭解與釐清交互作用對於依變項的影響大小與情形。

二、近年來，測驗學者及校務研究 (Institutional Research) 透過資料採礦 (Data Mining) 來分析大數據或資料庫資料，以瞭解測驗變項之可能分配的組型、群組或關聯性是重要的趨勢。資料採礦包括監督式 (supervised) 及非監督式 (unsupervised) 取向。請問，這兩種取向的意義與差別、資料採礦特性、兩種取向各有那些技術可以應用。(25分)

試題評析

本題是涉及資料採礦的命題，雖然與統計學有關，也與當今大數據分析的趨勢相符，但是偏重資料收集的技術性與方法論，與高階公務人員考試之行政業務導向卻背道而馳，實屬冷門且艱澀的考題。

答：

(一)資料採礦可分為監督式及非監督式兩種取向：

- 1.監督式：是找出特定變數的數值有何意義，並選定輸出變數，讓電腦找出如何對其進行評估、分類或是預測。也就是嘗試對已知的型態關係進行解釋。如選定「可能會倒帳的客戶」為輸出，將變數(如年收入及性別等)放入進行分析。如MBR。
- 2.非監督式：則沒有輸出變數，此模式的目的是在於讓電腦自行找出顯著關聯的型態。如購物籃分析。

(二)兩種取向的差異如下：

- 1.監督式資料採礦是希望能夠解答特定問題時使用。其程序包含以下四步驟：

(1)確認原始資料來源

最佳的資料來源是既存的資料倉儲，倉儲內部的資料皆經過固定格式的整理，方便進行分析。然而多數的情形下企業並沒有建立資料倉儲，必須收集各部門的相關資訊。

(2)彙整資料以供分析

如同假設檢定時的資料彙整原則，加入有用的變數可以增加產出有用結論的機率。許多資料採礦的情形下，資料都必須被分為「訓練集」(建立初始模型)、「測試集」(調整初始模型)及「驗證集」(衡量模型表現)。資料量在資料採礦中通常越多越好，而變數也是可以儘可能的將已知的變數放入模型，可能會有意想不到的結果。

(3)建立並測試分析模型

知識發掘的方法在建立模型的過程中容易出現過度學習的情形，也就是模型可能陷入只能解釋在訓練集樣本的關聯，而沒辦法一體適用。此時就要靠測試集將模型修正成較能通用的型態。

(4)評估分析模型

利用驗證集的樣本來評估模型分析的錯誤率，來決定分析模型是否有用。

- 2.非監督式知識發掘常常是進一步監督式知識發掘的前置作業，其步驟如同監督式，但是又多了另外兩個步驟：

(1)確認監督式知識發掘的可能目標。

(2)產生新假設並檢定。

(三)可以使用的技術有：

- 1.關聯法則
- 2.資料分類
- 3.預測分析
- 4.群集分析
- 5.異常分析
- 6.趨勢分析

7. 決策樹
8. 單純貝氏分類器

【參考書目】

謝邦昌、鄭宇庭(2016)，《資料採礦之技術及應用》，新陸書局。

三、S-P 表 (Student-Problem Chart) 分析是近年來測驗評量很重要的實務議題，請說明 S-P 表的意涵與特點，(10 分) 以及使用注意事項。(15 分)

試題評析	本題針對S-P表的命題屬於心理測驗學的範圍。主要應與反應組型以及注意與差異等幾個係數連結加以陳述，不困難但是偏小範圍的題目。
考點命中	《高點·高上107年高普考題神》教育測驗與統計，傅立葉編撰，頁5，解釋名詞(九)。

答：

(一)S-P表所關心的課題是學生在測驗試題上的作答「反應組型」，嘗試以幾個指標化數據做為診斷或判讀該反應組型是否為不尋常(unusual)或異常(aberrant)的一種測驗分析方法。這種分析測驗資料的方法，特別適用於以班級為單位的少數人資料的分析，尤其更適合用於形成性評量的測驗資料分析，它是一種屬於不對母群體特性設定任何假設值的統計推論方法-無母數統計方法(nonparametric method)。

(二)S-P表的使用注意事項有以下幾點：

- 1.S-P表的編製與分析之主要目的，在於獲取學習輔導有關的線索或訊息，以作為診斷之用。
- 2.S-P表的分析資料不是絕對的，教師在應用S-P表分析結果進行判斷的參考時，亦應參考其他資料，予以綜合的判斷和應用。
- 3.S-P表的正確判讀，端賴對學生和試題均熟稔的任課教師才能勝任，其他教師無能為力，除非他們也熟悉該班學生和該份試題。因此，S-P表對培養教師的教學能力，具有莫大的幫助。
- 4.用在形成性評量資料之S-P表，不宜作為排定學生名次之用途。
- 5.不宜拘泥於注意係數的大小。所謂注意係數較大時，不外是提醒：「還有其他問題，或是學生的反應組型相較於全體的反應組型有所差異，宜加以檢討」而已，數值本身並無法提供具體原因的說明。因此，教師仍應追根究底，方是根本解決之道。
- 6.注意係數之值為0.50，只是提醒要注意或檢討的信號而已。
- 7.當試題數或學生數過少時，可以不必計算注意係數，即使計算出來，亦不宜使用，因為它們的誤差極大，不具有代表性和正確性。
- 8.不同S-P表間的通知係數之比較，並沒有意義。只有在同一個S-P表中之注意係數值，才有比較的意義。
- 9.S曲線與P曲線相接近(即差異係數小)時，未必是一種好現象，亦應加以檢討。

四、解釋名詞：(每小題 5 分，共 25 分)

1. 測量標準誤 (Standard Error of Measurement)
2. 陸軍普通分類測驗 (Army General Classification Test, AGCT)
3. 測驗等化 (Test Equating)
4. 定錨試題 (Anchor Items)
5. 高斯分配 (Gaussian Distribution)

試題評析	本題的五個解釋名詞除了高斯分配外，都是心理測驗的範圍。除了定錨試題外，其他較為簡單。
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，頁15-16。 2.《高點·高上教育測驗與統計講義》第一回，傅立葉編撰，頁35-36。

答：

(一)測量標準誤

在一個團體的測驗分數中，全體誤差分數的標準差估計值，稱為測量標準誤(Standard error of measurement, SE_{meas})。亦即在多次測量後，每次實得分數與真正分數的差(測量誤差)的分配的標準誤。可

以做為信度計算之投入因子，也是個別受試者真正分數信賴區間估計的重要資訊。

(二)陸軍普通分類測驗

為美國陸軍在第二次世界大戰期間，因急需篩選大批新兵投入戰場，而由心理測驗專家學者組成委員會，負責設計一套團體智力測驗，稱為陸軍普通分類測驗。此測驗包括語文理解、算術推理與方塊計算等三個分測驗。使用平均數為 100，標準差為 20 的標準分數常模。

(三)測驗等化

所謂測驗等化(equating)是一種依據教育評量理論而發展的統計方法。由於不同試卷間存在難易度、信度等方面的差異，故需採行此方法以維持測驗結果評斷之公平性。測驗等化之目的在於使每次測驗分數間可客觀、有效進行比較，以確定每次各科測驗合格之考生能力標準維持一致。目前等化技術已被廣泛採用於國內外各項專業證照、語文與學力測驗中。

(四)定錨試題

錨定效應(Anchoring Effect，或 focalism)是一個心理學名詞，乃認知偏差的一種。當人類在進行決策時，會過度偏重最早取得的第一筆資訊(這稱為錨點)，即使這個資訊與這項決定明顯無關。也就是說，在進行決策時，人類傾向於利用最早取得的片段資訊，以快速做出決定，然後在接下來的決定中，再用第一個決定為基準點，逐步修正。但是人類容易過度利用第一個錨點，來對其他資訊與決定做出詮釋，當第一個參考用的錨點與實際上的事實之間的有很大出入，就會造成偏誤。特別是在試題反應理論的假設下進行等化設計時，不同的測驗中，必須包含一部分的定錨試題，以便作為測驗間的連結之用。

(五)高斯分配

常態分配(normal distribution)因為主要是由學者高斯所提出，所以又稱高斯分配。高斯分配之所以重要是因為其機率密度函數的圖形為鐘形曲線，再加上對稱性，使得很適合做為不少母體之機率模式。更重要的原因是由於在中央極限定理(Central Limit Theorem)下，常態分佈常可做為不少大樣本的近似分配以進行後續推論統計的應用與解釋。

高上

【版權所有，重製必究！】