

《教育測驗與統計》

試題評析	今年考題典試委員的命題較過去幾年著重統計數學計算，尤其偏好重複測量的相依樣本設計之統計方法。第一題的統整綜合問答題與第三題的一因子變異數分析皆是，對於今年的考情扮演舉足輕重的腳色。另外兩題則以測驗學為主軸，不過難度不高，其中又以第二題的測驗差異分數的信賴區間估計可以看出考生的準備功夫，計算與答題上都需要費點心思。綜觀今年考題之特性與難度，一般程度考生應可以有65~75分的成績，程度較佳同學可能在80分左右。
------	---

一、試述實驗設計中對於重複測量的資料處理有那些常見的統計方法？(25分)

試題評析	本題涉及跨不同章節之重複測量的相依樣本設計的數個統計方法。考生必須非常熟悉且能舉例敘述這些統計方法的內容、適用資料類型、以及檢定方法與結論的解釋與應用。包括： t 、 z 、 χ^2 與 F 檢定。對於考生平日的投入準備與統整能力有相當大的考驗。難度高低因人而異。
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，第11章，頁25；第12章，頁35-36；第13章，頁48-50。 2.《高上104地方特考重點題神》最新之議題重點整理(二)推論統計，5.卡方統計法，頁4-5。

答：

實驗設計中對於重複測量的資料處理，常見的統計方法有以下幾種：

(一)相依樣本設計的平均數差異顯著性 t 檢定

利用重複測量的資料，進行平均數是否存在顯著差異的 t 檢定，可以依檢定結果結論受試者的前後表現是否進步，並進一步取得某教學措施是否有效的統計證據。此統計方法的適用資料類型為連續變數。

(二)相依樣本設計的比例差異顯著性 z 檢定

當研究著重於受試者對於某項議題或事件的看法是否因為特定活動而有前後不同的意見表態，則可以透過相依樣本設計的比例差異顯著性 z 檢定，進行在某一顯著水準下的假設檢定。此統計方法的適用資料類型為間斷型的類別變數。

(三)改變顯著性的卡方 χ^2 檢定

概念類似前述(二)的比例差異顯著性 z 檢定，而且兩者之間存在著檢定統計量的一個特殊關係，也就是

$$(z)^2 = \chi^2$$

當拒絕未存在顯著改變的虛無假設，檢定結論為受試者在(事件)前後的態度發生顯著改變。

(四)受試者內設計的變異數分析 F 檢定

當研究問題著重於探討同一群受試者接受重複測量後，其反應表現是否相同時，涉及屬於自變項的實驗處理(分組)以及依變項的反應變項之間的因果關係探討，則為受試者內設計的變異數分析統計方法的應用時機。當自變項只有一個時，則屬一因子變異數分析的 F 檢定。如果檢定結果拒絕平均數全等的虛無假設，則可以進一步探討造成顯著不全等的組別差異為何。此統計方法的適用資料類型，自變項為間斷型的類別變數，依變項為連續變數。

二、假定性向測驗的分數是以 T 分數表示，其中語文推理測驗的信度為0.85，數學性向測驗為0.90。

某生在性向測驗上得到語文推理性向分數為54分，數學性向分數為62分，兩性向分數相差8分，

試問該生的數學性向是否優於語文性向？(25分)

試題評析	本題涉及標準分數之一的 T 分數與重測信度的計算應用。另外，最大的答題挑戰是兩測驗分數是否存在優劣差異，必須先計算兩測驗差異分數的標準誤，然後自行設定較常用的信賴水準，再利用兩性向測驗分數差異之信賴區間估計結果是否包含0進行結論。
考點命中	《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，第15章，頁15。

答：

依題意，由於性向測驗的分數以T分數表達，某生在該性向測驗之語文推理與數學性向測驗分數分別為54分與62分，雖然利用T分數的常模可以得出其標準化z分數分別為：

$$54 = 50 + 10z, z = 0.4$$

$$62 = 50 + 10z, z = 1.2$$

如此看來，似乎該生的數學性向優於語文性向。但是，為了排除誤差所造成的影響，應利用題目提供的信度資訊進行進一步的檢視與探討如下：

首先，計算兩測驗差異分數的標準誤：

$$\begin{aligned} SE_{diff} &= s\sqrt{2 - r_{XX} - r_{YY}} \\ &= 10\sqrt{2 - 0.85 - 0.9} \\ &= 5 \end{aligned}$$

接著，計算兩測驗差異分數的95%信賴區間

$$(62 - 54) \pm 1.96(5) = 8 \pm 9.8 = (-1.8, 17.8)$$

由於上述信賴區間包含0，表示兩性向測驗的差異分數在考量誤差成分後，並未存在顯著差異。因此，該生的數學性向並未顯著優於語文能力。

三、下表為8名受試者在四種色光的反應時間，試問受試者在四種色光的反應時間是否有所不同？

($F_{95(3,21)} = 3.07$) (25分)

受試者	紅光	橙光	黃光	綠光	ΣS	ΣS^2
A	4	2	4	6	16	72
B	6	4	6	6	22	124
C	2	2	4	4	12	40
D	4	4	4	6	18	84
E	2	2	4	4	12	40
F	2	4	4	4	14	52
G	2	2	2	2	8	16
H	2	4	4	4	14	52
ΣX	24	24	32	36	116	
ΣX^2	88	80	136	176		480

試題評析	本題考的是重複測度下之相依樣本設計的一因子變異數分析。雖屬冷門的計算題，惟本班考生平日仍有計算練習且願意投入者，在解題上應無太大阻礙與困難。無論如何，也是今年考試決定勝負的關鍵之一。
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13章，頁49-50。

答：

依題意，8名受試者在4種色光的反應時間資料屬於重複測度的相依樣本設計；由於要回答受試者在四種色光的反應時間是否有所不同，正確的統計方法為一因子變異數分析，茲進行計算與假設檢定如下：

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: \mu_j \text{不全等}$$

利用題目提供的資訊與統計量，可以進一步計算：

$$SS_t = \sum \sum x_{ij}^2 - \frac{(\sum \sum x_{ij})^2}{N} = 480 - \frac{116^2}{32} = 59.5$$

$$SS_{b.subject} = \frac{\sum S^2}{k} - \frac{(\sum \sum x_{ij})^2}{N} = \frac{1808}{4} - \frac{116^2}{32} = 31.5$$

$$SS_{w.subject} = SS_t - SS_{b.subject} = 59.5 - 31.5 = 28$$

$$SS_{b.treatment} = \frac{\sum (\sum x_j)^2}{n} - \frac{(\sum \sum x_{ij})^2}{N} = \frac{3472}{8} - \frac{116^2}{32} = 13.5$$

$$SS_{residual} = SS_{w.subject} - SS_{b.treatment} = 28 - 13.5 = 14.5$$

因此，檢定統計量為

$$F = \frac{13.5}{14.5} = 0.931$$

而題目提供臨界值 $F_{.95(3,21)} = 3.07$

由於0.931的檢定統計量小於3.07的臨界值，不拒絕虛無假設，利用這8名受試者的樣本資料，在0.05的顯著水準下，無法提供充分證據足以結論8名受試者在4種色光的反應時間有所不同。

四、古典測驗理論以真實分數模式為主，其理論模式的發展為時已久。試從試題反應理論的觀點，申論其缺點。(25分)

試題評析	本題主考古典測驗理論因其時空背景的部分簡化假設，在後續較成熟的試題反應理論觀點看來有一些缺點需要被闡明。題目簡單，應為今年試題中最容易的一題。
考點命中	《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，第19章，頁43。

答：

從試題反應理論的觀點，古典測驗理論的缺點可以分別敘述如下：

(一)採用之指標(難度、鑑別度、信度)為樣本依賴之指標：

實務上，難度和鑑別度都不是不變值，它們會隨著考生群體的能力分配的不同，而有不同的估計值出現，這些估計值都是樣本依賴(sample dependent)的估計值。因此，用來決定試題指標的樣本組能否適切代表測驗所要測量的母群體，便成為決定某種審慎選擇試題的技術能否成功的主要因素。

(二)以一相同之測量標準誤作為每人之測量誤差指標：

每一位受試者的測驗誤差應有差異，用同樣的測量標準誤進行後續有關的真實分數信賴區間估計實不合理。

(三)無法對於非複本但功能相同之測驗提供有意義的比較：

由於對信度的假設是建立在複本測量之概念假設上，然實務上能夠施測複本的主客觀條件不多，因此所謂有意義的比較缺乏立論依據。

(四)忽視受試者之試題反應組型：

認為原始得分相同的受試者，其能力必定一樣；其實不然，即使原始得分相同的受試者，其反應組型也不見得完全一致。因此，其能力估計值應該會有所不同。

【版權所有，重製必究！】