

# 《教育測驗與統計》

## 試題評析

第一題：本題為推論統計中，兩個母體參數差異之假設考驗。命題目的在要求考生，熟悉獨立樣本t考驗的參數為平均數，並能寫出該考驗方法得以順利進行之假定條件。題目簡單易答，本班考生應能完全掌握此題分數。

第二題：本題出題方向以極端值對於統計量數之平均數與標準差的影響。兩個公式對於本班同學而言，是再熟悉不過的，應能掌握所有分數。

第三題：本題將信度相關重要的計算作為答題重心，考的是考生平日準備的完整度。本班教材清楚列示考題中之斯布公式與盧隆公式；授課時也仔細說明與練習斯布公式的計算，以及alpha係數的公式意涵。

第四題：本題相當意外的針對非測驗主流的標準參照測驗的難度與鑑別度指標之試題分析命題。平日授課已經提及PPDI指標。如果同學熟悉標準參照測驗的特性，則可以容易的舉一反三，寫出另外兩個鑑別度分析指標。至於難度分析的特性則相對容易。

第五題：本題考的是複迴歸分析的模式檢測問題。本班教材第一回第八章對於複迴歸分析著重在共線性的成因、影響與修正等議題，事實上，這題雖然並未超出考選部界定本科的考試範圍，但是偏重理論的內容卻較偏向研究所的考試命題方向。

整體言之，一般程度的考生應能在第一題到第三題充分掌握分數。第四題與第五題則較有鑑別力，能有75分以上的成績已經相當有水準！

一、請舉一個資料蒐集設計來說明適用獨立樣本t考驗的研究問題及基本假定，並列出其虛無假設( $H_0$ )與對立假設( $H_1$ )。(25分)

答：

(一)一位國二老師想了解「其任教學校國中二年級生的數學能力是否在男女生之間存在顯著差異」，在此一研究問題下，該老師進行該校國二學生男女生各25名，該學期第二次段考數學科成績的隨機抽樣。由於男生與女生為獨立母體，兩性別數學成績差異之檢定參數為兩母體平均數差異( $\mu_1 - \mu_2$ )，因此虛無假設與對立假設為：

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

(二)基本假定為

1. 小樣本( $n < 30$ )
2. 兩個母體標準差皆為未知
3. 透過F考驗，得知兩個母體變異數同質性的假設不被違反因而適用獨立樣本t考驗。

## 【高分閱讀】

傅立葉，《教育測驗與統計》，第二回，頁18~20。

二、某項入學考試有50位考生報考，其中48位考生報名一般組，年齡介於20到30歲之間，中位數是25歲，年齡分布呈常態分配。另外有2位考生報名在職組，年齡分別是45與55歲。請問：(20分)

(一)這2位在職組考生對於全體50名考生的平均年齡與標準差的計算有何影響？(請列舉統計公式來說明你的答案)

(二)50名考生年齡平均值的位置是否為PR=50？為什麼？

**答：**

(一)由於2位在職組考生的年齡45與55歲，相較於其他48位一般組考生的中位數25歲而言，因為幾乎是其一倍的年齡，已屬於極端值，因此在計算全體50位考生年齡之平均數與標準差時，將導致兩個統計量數的錯估，進而影響對於考生年齡特性的解釋。

以平均數的計算公式而言， $\mu = \frac{\sum_{i=1}^{50} x_i}{50}$ ，因為分子部分為所有考生年齡的加總，計算結果將高估全體考生的平均年齡。

同理，標準差的計算公式， $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$ ，其分子的離均差平方和，因2位在職生與全體考生平均年齡的差異過大而擴大，也將導致全體考生年齡標準差的高估，而錯誤解讀考生年齡的差異較大。

(二)由於高估的平均年齡，50名考生年齡平均值的位置，將不會是  $PR = 50$ 。而會是比50的PR值更大的位置。

**【高分閱讀】**

傅立葉，《教育測驗與統計》，第一回，頁10~11，14。

三、假設有一個測驗是由四個題目所組成的，各題的平均數、標準差與變異數，以及折半之後的各種有關數據分別呈現於下表中，如果  $H_1$  與  $H_2$  之間的相關為.9154。請根據這些數據來回答下列的問題：(每小題5分，共25分)

某份測驗資料各種折半法之基本統計表

變項	$\mu$	$\sigma$	$\sigma^2$
各題目			
I1	3.00	1.33	1.76
I2	4.00	1.10	1.20
I3	3.40	1.36	1.84
I4	3.80	1.47	2.16
各種折半測驗			
H1 (I1 + I2)	7.20	2.40	5.76
H2 (I3 + I4)	7.20	2.71	7.36
Total	14.40	5.00	25.04
三種折半法之差異分數			
D(H1- H2)	0.00	1.10	1.20

(表內數字以四捨五入取至小數點第二位)

(一)請計算經過Spearman-Brown公式校正後的折半信度係數？

(二)當用Spearman-Brown法來估計折半信度係數時，有一個重要的假定(assumption)，請說明該假定為何？

(三)請計算Rulon折半信度係數？

(四)請計算alpha係數？

(五)一般的教科書都提到alpha係數的值等於所有可能的折半係數的平均數，這裡所說的折半係數是指用Spearman-Brown校正得到的信度係數或是用Rulon的方法所計算得到的？

**答：**

(一)由  $H_1$  與  $H_2$  之間.9154的相關係數可知，該測驗的折半信度係數正是.9154。計算斯布校正公式，取得還原為整個測驗的信度係數如下：

$$r = \frac{2(.9154)}{1 + (2-1)(.9154)} = .975$$

(二)當用Spearman-Brown法來估計折半信度係數時，有一個重要的假定是兩半測驗分數的變異性相等。也就是說，兩個半測驗在折半時應盡量做到難度、題數、內容、形式等各方面的一致，才能夠在校正後獲得最高且正確的信度係數。兩半測驗分數的變異數相等，亦即要滿足變異數同質性的假設，若違反這個假設，將導致高估測驗的信度(Cronbach, 1951)，所估計出來的信度係數值將會比其它內部一致性方法所估計出的信度係數值還高。

(三)Rulon折半信度係數的校正計算如下：

$$r_{XX(R)} = 1 - \frac{1.2}{25.04} = 1 - 0.0479 = 0.9521$$

(四)alpha係數計算如下：

$$\alpha = \left(\frac{4}{4-1}\right)\left(1 - \frac{1.76+1.2+1.84+2.16}{25.04}\right) = 1.333(0.278) = 0.371$$

(五)Crocker和Algina(1986)指出： $\alpha$ 係數是估計信度的最低限度(lower bound)，是為所有可能的折半係數之平均數。欲估計內部一致性係數， $\alpha$ 係數優於折半法。Cortina(1993)歸納國外有關 $\alpha$ 係數研究的觀點，發現其具備以下四項特性：

1.  $\alpha$  係數是所有可能的折半係數之平均數；
2.  $\alpha$  係數是估計信度的最低限度；
3.  $\alpha$  係數是表示第一個因素飽和度的量數(a measure of first-factor saturation)；
4. 當試題計分為二分名義變數時(答案登錄為1或0)，則  $\alpha$  係數之值與KR20之值是相同的。

當測驗題目之同質性高，則 $\alpha$ 係數與折半信度所估計的值近似，若題目之異質性高，則 $\alpha$ 係數會低於折半信度所估計的值，故 $\alpha$ 係數常被稱為估計信度的最低限度(傅粹馨，民91)。基於此，「alpha係數的值等於所有可能的折半係數的平均數」，這裡所說的折半係數是指用Spearman-Brown校正得到的信度係數。

**【高分閱讀】**

傅立葉，《教育測驗與統計》，第三回，頁12。

四、所謂標準參照測驗，一則清楚描述測驗所欲評量的內容領域，二則清晰描繪出所欲測量的建構，故其試題分析目的在檢驗試題是否能評量出學生在測驗表現所展現的所知與所能，已達到精熟水準或既定水準(通常以決斷分數區精熟與否)。當教師使用標準參照測驗時，大多希望其教學是成功有效的，故此種測驗的試題分析目的在回答「試題能否測量到教學的效果」。

(一)試說明三種標準參照測驗所使用的鑑別度分析法。(15分)

(二)請簡單描述標準參照測驗的難度分析有何特色?(5分)

**答：**

(一)標準參照測驗所使用的試題通常會比常模參照測驗所使用者容易些，因為教師多半期望大多數的學生在測驗上都有良好的表現，獲得好成績。在該類測驗上，教師若仍沿用常模參照測驗的試題分析方法，則高分組與低分組學生在試題上的答對百分比值都會接近1，鑑別度指標值則將趨近於0。由此可見，標準參照測驗的試題分析方法應不同於常模參照的試題分析法，方能避免試題分析產生無法解釋的效果。

標準參照測驗使用的鑑別度分析法可計算以下三種指標：

1. 教學敏感度指標

$$PPDI = P_{\text{post}} - P_{\text{pre}}$$

為Cox和Vargas(1996)所提出之同一組受測者在教學前進行前測和教學後進行後測，並比較前、後測答對

百分比之差異的方法。正常教學敏感度指標值大約是介於.10到.60之間。

(1)D接近0，代表該試題的鑑別度越低，不能反映出預期的教學效果；

(2)D接近1，代表該試題的鑑別度越高，對教學效果的敏感度越大，代表該試題的品質最佳；

(3)D是負值，代表該試題是個反向作用題，反映出教學效果是有問題的，代表該試題品質極為不良，應予淘汰。

### 2.精熟組與非精熟組的答對百分比差異指標

$D = P_p - P_f$   $P_p$ ：精熟組的答對比率  $P_f$ ：非精熟組的答對比率

### 3.接受教學組與未接受教學組的答對百分比差異指標

尋找各方面條件(如智力、學業成績等)均相等的兩組學生為受試者，其中一組接受教學，另一組未接受教學，然後分別計算其難度指標值的差異。

第二與第三種指標的解釋與計算與第一種相似。

(二)一般來說，標準參照測驗的難度分析是沒有必要進行的，因為該類測驗的目的不同於常模參照測驗。標準參照測驗的難度分析有以下幾點特色：

1.標準參照測驗試題的難度指標，應和教學前預設的通過標準(精熟標準)一致。例如：假設以80%的試題答對率或通過分數為精熟標準時，則理想的難度指標應訂為.80左右。

2.教師對於該類測驗的試題選擇標準，是依據教材內容和實際教學情況，做為判斷精熟標準或難度指標。

3.解釋上和常模參照測驗不同：

(1)因希望學生在教學後能精熟教材，故在教學有效的前提下，多數题目的難度指數(通過百分比)均很大(難度指數反映的教學品質多於题目的品質)。

(2)標準參照測驗的難度應和事先設定的相同，例如：以85%為精熟標準時，理想的難度指數應訂在0.85左右。

#### 【高分閱讀】

傅立葉，《教育測驗與統計》，第三回，頁24-25。

### 五、請簡要說明如何偵測多元迴歸中預測變項(predictors)的共線性(multicollinearity)? (10分)

**答：**

多元迴歸模式中有關預測變項的共線性，指的是獨立變項之間的相依程度。共線性主要影響有：迴歸係數的正負值、迴歸係數的估計標準誤錯估、以及迴歸係數的推論統計等。其偵測方法有以下幾個：

對於單一預測變項共線性的檢視最常用的指標有：

#### (一)容忍度(Tolerance)

容忍度的探討，須先計算模式中其它變項總合來說，對此一變項(視為效標變項Y)的預測力--R平方(即決定係數)，然後再用一減去「R平方」而得，算式符號表示為 $1 - R_j^2$ 。其值愈大愈好，表示該預測變項有其不可取代的預測能力，不會是共線性變項。

#### (二)變異數膨脹係數(因子)(Variance inflation factor, VIF)

其值正好是上述容忍度的倒數，計算式為 $\frac{1}{1 - R_j^2}$ 。計算結果大於10的數值表示：第j個預測變項正是共線性變項。

對於整體迴歸模式共線性的偵測則使用以下兩個指標：

#### (三)特徵值(Eigenvalue)

#### (四)條件指標(Conditional index, CI)

#### 【高分閱讀】

傅立葉，《教育測驗與統計》，第一回，頁46-47。