

【教育行政】

《教育測驗與統計概要》

試題評析

試題簡單且基本，熟讀本班出版之相關書籍，定能獲取高分。

一、某研究者想比較不同的補救教學方法對克服英文學習困難的影響。該研究者設計三種教學方法，並以隨機分派方式將學生分派到三種教學方法（每組四名），所蒐集到的資料如表一所示。經計算後得知，組間離均差平方和（SSBG）為 86，組內離均差平方和（SSW）為 8.25。

（一）請完成表二的變異數分析摘要表的括弧（要寫計算過程）。（20 分）

（二）如何知道那兩組間有差異？請具體說明你可能採用的作法及理由為何（無須計算）？（5 分）

表一 受試者在三種方法的得分

方法一(n=4)	方法二(n=4)	方法三(n=4)
4	8	3
5	8	2
4	9	1
6	10	3

表二 變異數分析摘要表

變異來源	SS	df	MS	F
組間	86	()	()	()
組內	8.25	()	()	
全體	94.25	()		

答：

ANOVA 表：

SV (變異來源)	SS (平方和)	df (自由度)	MS (均方)	F-test
B (組間)	SSB	a-1	MSB	$\frac{MSB}{MSW}$
W (組內)	SSW	n-a	MSW	
Total	SST	n-1		

其中，SSB+SSW=SST

$$(a-1) + (n-a) = n-1$$

$$SSB \div (a-1) = MSB$$

$$SSW \div (n-a) = MSW$$

	(1)	(2)	(3)	
	4	8	3	
	5	8	2	
	4	9	1	
	6	10	3	
$\sum X$	19	35	9	$\sum \sum X = 63$
$\sum X^2$	93	309	23	$\sum \sum X^2 = 425$
$\bar{X}_{\cdot j}$	4.75	8.75	2.25	$\bar{X}_{\cdot\cdot} = 5.25$

計算：

(一) $H_0: \mu_1 = \mu_2 = \mu_3$

H_1 ：至少有二組以上不等

(二) $SST = \sum \sum (X_{ij} - \bar{X}_{\cdot\cdot})^2$

$$\begin{aligned}
 &= \sum \sum X_{ij}^2 - \frac{(X_{\cdot\cdot})^2}{n \cdot a} \\
 &= 425 - \frac{(63)^2}{4 \times 3} \\
 &= 425 - \frac{3969}{12} \\
 &= 425 - 330.75 \\
 &= 94.25
 \end{aligned}$$

$$\begin{aligned}
 SSB &= \sum \sum (\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2 \\
 &= \sum \frac{(X_{\cdot j})^2}{n} - \frac{(X_{\cdot\cdot})^2}{na} \\
 &= \left[\frac{(19)^2}{4} + \frac{(35)^2}{4} + \frac{(9)^2}{4} \right] - \frac{(63)^2}{4 \times 3} \\
 &= \frac{361 + 1225 + 81}{4} - 330.75 \\
 &= 416.75 - 330.75 \\
 &= 86
 \end{aligned}$$

$$\begin{aligned}
 SSW &= SST - SSB \\
 &= 94.25 - 86 \\
 &= 8.25
 \end{aligned}$$

$$(三) MSB = \frac{SSB}{df_B} = \frac{86}{3-1} = 43$$

$$MSW = \frac{SSW}{df_W} = \frac{8.25}{3(4-1)} = 0.91666$$

$$(四) F = \frac{MSB}{MSW} = \frac{43}{0.92} = 46.73$$

(五) 臨界值 $F_{0.05}(2,9)$ = 《本題題目未提供》

變異數分析摘要表

變異來源 (S V)	SS	df	MS	F
組間	86	2	43	46.73
組內	8.25	9	.92	
全體	94.25	11		

* $p < 0.05$

比較任何兩組是否有差異 => 多重比較：

(一) 杜凱氏 HSD (honestly significant difference) 考驗法適用於每組人數相等。杜凱氏 HSD 法和紐曼-柯爾法 (Newman-Kuels) 都是根據下列公式所示的「差距考驗」(studentized range test) 而來：

$$q = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{\sqrt{\frac{MSW}{n}}}$$

(二) 紐曼-柯爾 (Newman-Kuels) 法

第二種可以適用於每組人數相等的差距考驗叫做紐曼-柯爾法 (NewmanKeuls method)，簡稱 N-K 法。其最大特色在於：它是依平均數之大小次序使用不同的臨界 q 值。

(三) 薛費法

不管是 HSD 法或 N-K 法，均只適用於各組人數均同為 n 的時候，以及只適用於比較一對兩個平均數之間的差異的時候。如果各組人數不相等，或者想要進行複雜的比較，亦即每次比較包含兩個以上平均數之間的差異時，就需使用薛費法 (Scheffe method，簡稱 S 法)。

【參考資料：余敏賢，教育測驗與統計重點整理 p407~p411 或教育測驗與統計 p19-1~19-10 (高點文化出版)】

二、請說明卷宗評量 (或檔案評量，portfolio assessment) 的意義、目的、具體的作法和優缺點？
(25 分)

答：

(一) 意義

1. 多角度、多向度的評量歷程，包括一切有關人員、事項及資訊的蒐集。
2. 指每個學生都有一個卷宗，以個人為單位蒐集資料，綜合其各項表現，利用時間上的連續，傳達個人的學習歷程。

(二) 目的

1. 將學生在教室中的作品依時間蒐集、整理，作為將來成長的紀錄。
2. 要評量一般傳統評量做不到的學生合作與社會能力。
3. 要促進學生自省、自評、自我瞭解的能力。
4. 要讓關心者能夠瞭解學生在校的表現。
5. 希望培養學生自我負責、獨立的能力。
6. 希望結合教導、學習、與評量，使之相互修正。

(三) 優點

1. 重視學生個別的需要、能力和興趣。

2. 強調學生真實生活的表現。
3. 師生共同參與評量的過程。
4. 幫助別人了解自己學習的成就。

(四)缺點

1. 相當費時、費力。
2. 學生個別差異大、評分表設計不易。
3. 評分的信度受到質疑。

(五)具體的作法

1. 每一科目製作一份檔案或整體學習卷宗。
2. 先列目錄，要有目錄碼及頁碼，所有資料全部 A4 規格化。
3. 注意美工、創意及個人特色。
4. 卷宗依據評鑑規準及個人成長目標來組織。
5. 每一卷宗文件盡量有「標題」、「註解」和「省思評語」用以說明內容並就卷宗內容作反省性的思考。

三、測驗的信度常會受測量誤差大小影響。如果將測量誤差來源大致歸納為三類：時間取樣誤差、內容取樣誤差和評分者誤差。相對應於上述三種誤差來源的信度計算方法各有那些？(25 分)
答：

信度的大小與信度的估計方法之間有密切的關係，不同的信度估計方法，各有不同的測驗誤差來源如下表：

各種信度類型及誤差來源

誤差來源	信度類型	解決的主要問題
時間抽樣	再測信度	情境的影響： ① 相關內容樣本所得分數受到不同測驗情境的影響如何？ ② 在不同測量時間所得分數的穩定性如何？
時間抽樣與內容取樣	複本信度	① 無論使用的複本測驗或實施的情境怎樣，測驗的一致性如何？ ② 在不同測量時間所得分數的穩定性如何？
內容抽樣	複本信度 (同時實施)	不同內容抽樣的影響： ① 測驗分數在相同情境下，是否受不同內容抽樣的影響？ ② 兩份仔細配合的複本測驗是否相等、平行或可交互使用？
內容抽樣	折半信度	① 測驗分數在相同情境下，是否受不同內容抽樣的影響？ ② 複本形式的信度係數為多少？
內容抽樣與內容異質	K-R 信度 (20 號公式)	① 測驗分數在相同情境下，是否受不同內容抽樣的影響？ ② 測驗的同質性如何？ ③ 每一個試題的反應一致性如何？
內容抽樣與內容異質	K-R 信度 (21 號公式)	測驗分數在相同情境下，是否受不同內容抽樣的影響？
內容抽樣與內容異質	α 係數	測驗分數在相同情境下，是否受不同內容抽樣的影響？
評分者的誤差	評分者信度	不同評分者的影響： ① 如果使用不同評分者，評分差異的程度如何？ ② 測驗的客觀程度如何？ ③ 不同評分者所得的結果是否可替換？

【註】

$$KR_{20} = \frac{1}{n-1} \left[1 - \frac{\sum pq}{S_x^2} \right]$$

$$KR_{21} = \frac{1}{n-1} \left[1 - \frac{\bar{X}(n-\bar{X})}{nS_x^2} \right]$$

$$\sigma = \frac{n}{n-1} \left[1 - \frac{\sum S_i^2}{S_x^2} \right]$$

(一) 信度的估計方法：

最常使用的信度估計方法有四類，分別如下：

1. 再測方法(test-retest method)
2. 複本方法(parallel-forms method)
3. 內部一致性方法(internal-consistency method)
4. 評分者方法(scorer method)。

(二) 信度的種類：

1. 再測信度：

再測信度(test-retest reliability)是最常用也最容易理解的信度求法，它是指將同一個測驗在不同時間上重複施測於同一樣本，並求二次測驗之間的相關係數，以證明其分數的穩定性；因此，再測信度又稱為穩定係數(coefficient of stability)。

一般而言，兩次施測間隔時間愈短，再測信度愈高；時間間隔愈久，隨機誤差來源便愈多，信度也就愈低。

再測信度的誤差來源包括下列數種：

- (1) 受測者個人在兩次測驗之間的變化，如動機、健康、學習、練習等因素。
- (2) 兩次施測情境間的變化，如溫度、光線、計時或計分上的錯誤、意外的干擾等因素。至於內容或試題取樣的誤差，由於前後兩次測驗題目完全一致，所以不構成誤差來源。

2. 複本信度：

複本測驗(alternate forms、equivalent forms 或 parallel forms test)是指兩份在內容、型式、題數、難度、指導語說明、時間限制與例題等方面都相當，都用來測量相同潛在特質或屬性，但試題卻不相同的測驗。拿兩個複本測驗給同一群學生施測，再依據測量得到的分數求相關，即為複本信度(parallel forms reliability)。因此，複本信度愈高，即表示兩份測驗所測量到相同特質或內容的程度愈高，所使用的樣本試題愈具有所要測量內容範圍的代表性。

實施兩份複本測驗的方式有二：一為在同一個時間連續施測，另一為間隔一段時間施測。前者所得到的複本信度，可以反應出測驗工具內容所造成的誤差有多少(即試題的抽樣誤差)，但無法反出學生本身狀況所造成的誤差(即時間的抽樣誤差)，這種複本信度又稱為「等值係數」(coefficient of equivalence)；而由後者所得到的複本信度，不但可以反應出測驗內容所造成的誤差，而且也可以反應出學生本身狀況所造成的誤差，這種信度又稱為「穩定且等值係數」(coefficient of stability and equivalence)。由於穩定且等值係數是使用複本式的再測方法(test-retest method with equivalent forms)，可以同時反應試題抽樣與時間抽樣所產生的誤差，所以是估計測驗信度的最好方法。

3. 內部一致性信度

只根據一次測驗結果所估計的信度，這種信度便稱作「內部一致性信度」(internal-consistency reliability)。屬於這種信度估計方法者，比較常用的計有：折半方法、K-R 方法與 α 係數等三種，分述如下。

(1) 折半方法(split-half method)：

所謂折半方法，即是利用單獨一次測驗結果，以隨機方式將其分成兩半，再求出這兩半測驗結果間的相關，即為折半相關(split-half correlation)。在理論上，折半相關和同時間施測所得的複本信度是一樣的，都適用來說明測驗式題抽樣的適當程度，不過，因為折半方法只憑一次測驗結果來求兩半分數的相關，所以其信度係數又稱為折半信度(split-half reliability)。

(2) K-R 方法(Kuder-Richardson method)：

另一種用來估計測驗信度的方法，適用於二分法計分(dichotomous scoring)資料的試題者，為庫-李(Kuder & Richardson, 1937)所創的 K-R 方法。

(3) α 係數(coefficient alpha)：

上述 K-R 信度僅適用於非對即錯的二分法計分測驗，對於非屬二分法計分的測驗資料，則無法適用。

但在心理與教育測驗中，有不少測驗的計分是屬於多元計分的方式，尤其是態度測量，測驗學者常用「李克氏五點評定量表」(Likert's five-point rating scale)，亦即是答「非常滿意」者給 5 分，答「同意」者 4 分，答「無意見」者 3 分，答「不同意」者 2 分，答「非常不同意」者 1 分，類似這種計分方式的測驗，K-R 信度的估計公式就無法適用，此時，必須使用 Cronbach (1951) 所發明的 α 係數。

(4) 評分者信度：

所謂評分者信度 (scorer reliability) 是指不同評分者之間，對於同一組試卷 (作品) 所評結困的一致性。在求算此種信度時，可依評分者人數及評分式而採取不同的統計方法，選擇時可參見下表。

求評分者信度可用的統計方法

		評分者人數		
		二名	二名以上	
評分方式	名次法	斯皮爾曼等級 相關係數	肯德爾和諧係數	(等級資料)
	分數法	皮爾遜積差 相關係數	變異數分析 (霍易特(Hoyt)分析法)	(等距資料)

【參考資料：余敏賢，教育測驗與統計重點整理 p41~p41(高點文化出版)】

四、解釋名詞並比較差異：(每小題 5 分，共 25 分)

(一) 性向 (aptitude) 測驗與興趣 (interest) 測驗

答：

(一) 性向測驗：測量先天「天賦能力」的測驗。就是常言的潛能 (potentiality)，此種與生俱來的能力又可分成兩種能力：

1. 普通能力 (general ability)：是英國心理學者史皮曼 (C.Spearman) 所稱的 G 因素 (general factor)，心理學上的智力 (intelligence) 指的就是這種能力。智力即是指個人此時此地學習事物所具備的先天能力。
2. 特殊能力 (specific ability)：即是史皮曼所稱的 S 因素 (specific factor)，心理學上的性向 (aptitude) 指的就是這種能力。性向的涵義較為複雜。就廣義的觀點言之，凡是先天天賦的能力 (潛能) 都可稱為性向，此時性向的觀念包含了智力的概念；如從狹義的觀點來看，則性向是專指天賦的特殊能力，泛指未來學習事物所具備的能力。狹義觀點的性向又可分成普通性向 (general aptitude) 及特殊性向 (specific aptitude) 兩種。

(二) 興趣測驗：

1. 興趣 (interest) 的意義：指個人全神貫注於某一活動的內在傾向。個人對其所喜歡的對象、目標、技能、知識或其他活動等，往往會投注更多的時間與努力去從事，並從中獲得滿足。例如，對科學有興趣的學生會花費很多時間閱讀科學習籍並參加有關科學活動。
2. 興趣量表：就是測量個人喜歡或厭惡某項活動的傾向。興趣量表 (interest inventory) 對於測量興趣的基本假定為：
 - (1) 興趣不是天生的，是個人從事某項活動結果而學得的。
 - (2) 年幼兒童的興趣相當不穩定，大約 20 歲以後就漸趨穩定。
 - (3) 不同職業的人有共同喜愛與不喜愛的活動。
 - (4) 興趣的強度因人而異。
 - (5) 興趣會引發個人的行為動力。
 - (6) 在相同職業或工作領域的成功者，其職業興趣與組型大約相同。
 - (7) 不同職業的成功者，其職業興趣不同。
3. 興趣測量方法：

- (1)直接觀察。
- (2)直接詢問。
- (3)興趣測驗(interest test)。
- (4)興趣量表。

【參考資料：余敏賢，教育測驗與統計重點整理第 8 章(高點文化出版)】

(二) 表面效度 (surface validity) 與內容效度 (content validity)

答：

- (一)表面效度(surface validity)：指測驗給人的第一印象好像是測量什麼東西，而非指測驗事實上能測量到什麼東西。故表面效度絕不能代替客觀的效度，不可認為改進了表面效度即可增加客觀的效度，但表面效度對一個測驗來說，仍然是很重要的，因為其易使受試者獲得親切感而願意合作。
- (二)內容效度(content validity)是指某測驗的試題能否適切地測量到此測驗之編製目的上所要測量的內容或行為領域(content or behavior domain)。此又可稱為「取樣效度」(sampling validity)或「邏輯效度」(logical validity)，前者表示它可以被視為檢查測驗內容取樣的代表性方法，後者表示這類效度多於邏輯判斷，而非源自實證的統計數據。

【參考資料：余敏賢，教育測驗與統計重點整理 p64(高點文化出版)】

(三) 受試者內設計 (within subject design) 與受試者間設計 (between subject design)

答：

單因子變異數分析有二種設計方式：

- (一)受試者間設計(between subject design)－為獨立樣本設計，所有受試者被隨機分派到 K 個不同的組別，分別接受這一個自變數的 K 個實驗處理 (或 K 個水準) 中的一個實驗處理，然因各組受試者彼此之間毫無關係，故又可稱為「受試者間設計」或「完全隨機化設計」。
- (二)受試者內設計(within subject design)－為相依相本設計，又可稱為「受試者內設計」或「隨機化區組設計」。下列三種情況均屬相依樣本設計。
 - 1.重複量數：以同一組受試者重複接受 K 種處理，以觀察其表現的差異。
 - 2.配對組法：選擇 K 組受試者在某一與依變數有關的特質方面完全相同，控制此配對的條件完全相同而比較 K 組受試者在依變數上的表現。
 - 3.同胎法：K 個先天條件完全相同的受試者接受不同的處理，例如將同卵雙生子分別放在貧窮與富有的環境中長大，以比較其差異。

【參考資料：余敏賢，教育測驗與統計重點整理 p403~404(高點文化出版)】

(四) 中數 (median) 與眾數 (mode)

答：

- (一)中數(median)－集中量數的第二種便是「中位數」，簡稱「中數」，以 Md 來代表。計算中位數之前，要先處理次序變數資料，亦即要先找出團體中佔在最中間位置那一個人，然後再看看這一個人的分數是多少。假使我們根據某變數之大小次序，將團體中之每個人的分數加以排列，那麼佔最中間位置那一個人所得到的分數是為中位數。我們以此一個人的得分來作為代表值，代表此一團體在此一變數方面的集中情形。
- (二)眾數－眾數也是集中量數之一，可用來代表一個團體的集中情形。因為眾數的英文為 Mode，故常以 Mo 來代表。眾數是指出現次數最多的數值，或最多人所得到的分數。

【參考資料：林清山，心理與教育統計學】

(五) 預測變項 (predictor) 與效標變項 (criteria)

答：

統計學中迴歸分析中所使用的名稱，例如 $Y=b*X+c$ ，b 為迴歸係數，c 為常數，X 為「預測變項」，而 Y 為「效標變項」。方程式中藉由預測變項 X 可以預測效標變項 Y。即用來預測的變項稱為「預測變項」或「獨變項」，而被預測的變項稱為「效標變項」或「依變項」。

【參考資料：林清山，心理與教育統計學】