

104 年特種考試地方政府公務人員考試試題

等別：三等考試

類科：教育行政

科目：教育測驗與統計

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)

【擬答】

(一)重複測量

在實驗設計中，重複測量作法，在統計分析上可以採用受試者內設計，係指同一位受試者同時參與多個實驗水準，換句話說，在進行實驗處理的過程中，研究者多次對同一位受試者在不同實驗情境，蒐集相關資料，例如：發展心理學常使用的縱貫法，便是針對同一群受試者在不同時間點施測，用以瞭解年齡對受試者的特質與能力之影響，像是 Thomas 和 Chess(1977)對 NY 的兒童進行氣質與父母親照顧方式的研究。

採用這樣的研究設計方式，針對研究資料之分析，有許多種作法，最簡單的分析作法為相依 t 檢定，假若涉及多組平均數則會使用變異數分析 (ANOVA)，若遇到有第三共同因素影響研究推論，則可以採共變數分析 (ANCOVA)，假若涉及時間趨勢的影響，可以採用時間序列分析，若涉及多個依變項，則採多變量分析 (MANOVA)。有此可以看出重複測量的作法在不同研究領域有很高的適用性，像是心理、教育、藥理學、經濟學等。

(二)以重複測量在 ANOVA 為例

1. 線性模式

第 j 組第 i 個人之分數

↑

$$X_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij} \rightarrow \text{殘餘誤差}$$

↓ ↓

總平均數 個別差異量

α_j 為實驗操弄效果

2. ANOVA Table

SV	SS	df	MS	F
實驗效果 A	$n \sum_j^p (\bar{X}_j - \bar{X}_{..})^2$	(p-1)	A-X/p-1	MS_A/MS_{AS}
個別差異 S	$p \sum_i^n (\bar{X}_i - \bar{X}_{..})^2$	(n-1)	S-X/n-1	
殘差 AS	$\sum_j^p \sum_i^n (X_{ij} - \bar{X}_j - \bar{X}_i + \bar{X}_{..})^2$	(p-1)(n-1)	$\frac{AS - A - S + X}{(p-1)(n-1)}$	
總變異 Total	$\sum_j^p \sum_i^n (X_{ij} - \bar{X}_{..})^2$	pn-1		

3. 例子

在心理學研究中，知覺研究常常使用重複測量的作法，像是不同色光對受試者的反應時間影響。

(三)優、缺點

公職王歷屆試題 (104 地方政府特考)

- 1.優點：僅需要較少人數即可進行實驗、個別差異的控制可以透過統計分析處理
- 2.缺點：人數會有流失問題、會出有順序效果 (order effect、practice effect、carryover effect 等) 或是疲倦問題

參考資料：

Kirk experimental design 3e

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？

【擬答】

(一)公式

$$T=10Z+50\dots\dots SD=10\dots\dots(1)$$

$$SEM_{diff.} = \sqrt{(SEM_1)^2 + (SEM_2)^2} = SD\sqrt{2-r_{11}-r_{22}} \dots\dots(2)$$

(二)計算過程

依題意所給條件帶入(1)，(2)

$$SEM_{diff.} = \sqrt{(SEM_1)^2 + (SEM_2)^2} = 10\sqrt{2-0.85-0.9} = 5$$

在 95%信心水準下，差異分數至少 9.8 分以上才能推論二份測驗分數有差異。

(三)結論

二分測驗的實得分數相差為 8，在 95%的信心水準之下，二份測驗的差異結果是不顯著的。

參考資源：

郭生玉 教育與心理測驗、葛樹人 心理測驗

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	$\sum S$	$\sum S^2$
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
$\sum X$	24	24	32	36	116	
$\sum X^2$	88	80	136	176		480

【擬答】

(一)統計假設建立與分析方式種類

1. 統計假設

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_0$$

H_1 : 任二個平均數不相等

2. 分析方式

依據題意本題分析方式為單因子受試者內分析

(二) ANOVA table

Source	SS	d. f.	MS	F	Effect size
色光	13.5	3	4.5	6.517*	0.482
個別差異	392.5	7	56.07		
殘差	14.5	21	0.69		
Total	420.5	31			

*: $P < .05$

(三) 結論

1. 依分析結果，不同色光對人的反應時間影響是顯著的，透過 Bonferroni 的事後檢定，可以發現整體效果顯著是來自紅光與綠光之差異所造成，惟推翻虛無假設的決策可能犯 $\alpha = .05$ 之錯誤

2. 依據統計分析的效果量來看，研究結果在實務上應用的可信程度為中等 (Cohen, 1988)

參考資料：

林清山 心理與教育統計

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)

【擬答】

(一)真分數理論

古典測驗理論的內涵，主要是以真實分數模式為理論架構，即觀察分數等於真實分數與誤差分數之和 ($X=T+E$)。其理論模式的發展已為時甚久，且發展得相當規模，所採用的公式簡單明瞭、淺顯易懂，適用於大多數的教育與心理測驗資料，以及社會科學資料的分析，為目前測驗學界使用與流通最廣的理論依據。

1. 古典測驗理論的假設

古典真分數理論是描述測量誤差如何影響觀察分數的一個簡單且很有用的模式。以下先列出古典真分數理論的假定，並逐一詳加解釋：

Allen & Yen (2002) 認為古典真分數理論有下列七個假設。

(1) $X = T + E$

此假設說觀察分數 X 可以分為兩部份，即 T 及 E ， T 是真分數， E 是誤差分數 (或測量誤差)。在古典真分數理論中，假定真分數與誤差分數具有可相加的關係 (而不是其他的關係，如乘法的關係)。

(2) $\varepsilon(X) = T$

此假設是說觀察分數的期望值 (母群平均數) 是真分數。此假定真分數的定義： T 是分數的理論分配之平均數，而分數的理論分配是重複對同一受試者實施同一測驗所得測驗分數之分配。

(3) $\rho_{ET} = 0$

此假設對於進一步的推演至為重要。其意為「誤差分數與真分數是沒有相關的」。也

就是說受試者的真分數高低，與其測量誤差沒有關係。

(4) $\rho_{E_1 E_2} = 0$

此假設 E_1 是測驗 1 的誤差分數，而 E_2 是測驗 2 的誤差分數。此假設是在說明「兩個不同測驗的誤差分數是無關的」。但如果測驗分數受疲勞、練習效果、作答時的心情或環境因素之影響很大時，則此假定就不合理。

(5) $\rho_{E_1 T_2} = 0$

此假設是在說測驗 1 的誤差分數 E_1 與測驗 2 的真分數 T_2 無關。

假定(1)至(5)提出古典真分數理論對測量誤差之定義。測量誤差是一個受試者的觀察分數非系統的或隨機的偏離其理論上的期望觀察分數之值。在古典測驗理論，系統的誤差不稱為測量誤差。

(6) 若二測驗之真分數相等 $T = T'$ ，誤差的變異數 $\sigma_E^2 = \sigma_{E'}^2$ ，則二測驗互為平行測驗。

如果兩個測驗的觀察分數 (observed scores) 分別為 X 及 X' ，此二測驗滿足上述(1)至(5)的五個假定，而且對每一個受試群體而言， $T = T'$ 以及 $\sigma_E^2 = \sigma_{E'}^2$ ，則此二測驗稱為「平行測驗」(parallel tests)，或是「嚴格的平行測驗」(strictly parallel test) (Lord, 1980)。

(7) 如果兩個測驗的觀察分數分別為 X_1 及 X_2 ，此二測驗滿足上述(一)至(五)的五個假定，而且對每一個受試群體而言， $T_1 = T_2 + c_{12}$ (c_{12} 是一個常數)，則此二測驗稱為「本質上 τ 等值測驗」(essentially τ -equivalent tests)。

在上述模式的各個假定中， X 是受試者的觀察分數， T 是受試者的真分數 (true score)， E 是誤差分數 (error score)。

但從古典測驗理論的假定中，我們知道此理論所有的假設均未限定觀察分數、真分數或誤差分數之分佈型態，也就是說無有關這些分數分佈型態之訊息。因此，除了上述古典測驗理論的假定之外，我們再假定：(1)測量誤差是獨立分配的，(2)測量誤差是常態分配的，其平均數為 0，變異數是一常數，(3)真分數對觀察分數之迴歸是直線的，亦即真分數和觀察分數二者均為常態分配 (Lord & Novick, 1968)。

(二) 依試題反應理論評論真分數理論的缺點

古典測驗理論一直是各種心理測驗編製與解釋之主要依據，但是由於此種理論沒有較強的假定，因此它的功能自然也受到相當程度的限制。古典測驗理論主要的缺點如下(王寶壙，1995)：

1. 抽樣變動大：試題參數值 (試題難度和試題鑑別度) 會因受試群體的抽樣變動影響甚大。
2. 能力難比較：受試能力參數會因受試題特徵之影響，因此，受試者必須接受相同或平行 (parallel) 的測驗題目，才能比較所測量能力之差異。
3. 缺乏預測力：沒有提供一個受試者可能答對個別試題的概率，當測驗設計者希望預測一個或多個受試群體測驗分數的特徵，或是要為某些受試群體設計某種特性的測驗時，此種訊息是必要的。
4. 複本難實施：平行的測量很難達到，受試者很難在兩次測驗上得到完全相同的結果。

古典測驗理論除了有上述四個缺點之外，它對很多測驗問題，如測驗設計 (test design)、測驗分數等化 (test score equating) 和試題偏誤 (item bias) 等問題，均無法提供滿意的解答。基於這些因素，心理計量學者乃發展出更適宜的心理測量理論。試題反應理論 (item response theory) 乃應運而生。

參考資料

郭生玉 教育與心理測驗、葛樹人 心理測驗、王寶壙 項目反應理論