

99 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、隨著資訊科技的發展，目前已有很多測驗利用電腦施測，其中適性測驗(adaptive test)是一種較受重視的施測方式，試就適性測驗回答下列問題：

(一)適性測驗係依據何種理論？

(二)適性測驗的基本原理為何？

(三)適性測驗何以必須要用電腦施測？

【擬答】：

(一)1. 適性測驗意義：「電腦適性測驗」(Computerized Adaptive Testing, 簡稱 CAT) 係結合項目反應理論、適性測驗、電腦科技而形成的智慧型電腦測驗，其特點在針對不同能力考生能適時適地出現合適考生題目，根據考生能力高低決定，每個試題對考生能力均有估計功能，不需大量試題即可在短時間精確測得考生能力，達到「因材施教」、「經濟有效」、「誤差最小」理想境界，簡稱「合身測驗」或「適性測驗」。

2. 適性測驗依據理論：項目反應理論(Item Response Theory)，指受試者面對試題進行作答的反應。受試者接受一組測驗施測後，項目反應理論即能反應隱藏在受試者內真實能力，也即為當初測驗設計者所想要測量之特質與能力。因此 IRT 有二個基本假設：1. 受試者在測驗項目的表現可由一組能力作預測與解釋。2. 受試者在測驗項目的表現與隱藏在受試者內真實能力關係可以用一個單調遞增函數描述(即項目特性函數 Item Characteristic Curve, 簡稱 I.C.C)；ICC 曲線表示，當能力水準 θ 增加時，其反應出回答正確的機率也將增加，即受試者某種潛在特質的程度與其在某一試題上正確反應的機率。又稱試題反應理論、潛在特質理論(latent trait theory, 簡稱 LTT)。

(二)適性測驗基本原理：針對受試者進行一系列測驗項目，根據受試者 2 至 3 題項目反應，得到此受試者「能力初步估計值」，再依此一估計由電腦中選取一組「提供受試者能力最多訊息」之項目施測，重覆上述步驟，直到已達預期測量精確度水準，即中止 CAT 測驗過程。

(應立志，民 83)

(三)適性測驗何以必須要用電腦施測：

1. IRT 許多特點正提供適性測驗所需理論基礎，考生不必作答同一份測驗也可比較其能力值。IRT 之項目訊息函數(Item Information Function)可提供適性測驗選取項目之用，而每個項目對測驗訊息是可加性性質，亦為 IRT 的特性之一。

2. 受測同時可利用電腦蒐集有用訊息，如考生每題作答時間，有助於瞭解受測者能力與觀察試題難易。

二、任何一種工具，在善於使用它的人手中，可以發揮其正面的、積極性的功能，產生有益於人的效果。若是落入不善於使用它的人手中，就常會發生誤用和濫用的情形，形成有害的後果；測驗工具亦復如是。因此測驗必須是由具備適當條件的人，才能使用。不具備是項資格者，無從干擾測驗之推展，關於測驗使用者的資格，應有測驗倫理的規範，試從下列三方面說明測驗使用者的資格應有那些規範？

(一)使用測驗的必要性。

(二)選用適當的測驗。

(三)測驗的發行。

【擬答】：葛樹人(民 92)提出心理測驗為有用科學程序和工具，其真正功能發揮仍賴測驗者正確使用，不良使用會導致個人與團體負面後果。測驗使用者應具有適當專業知識與訓練、專業精神、高度道德感與責任感，其目的不僅確定測驗工作者具有使用測驗所需之能力，尚可對測驗結果慎重評分，避免測驗誤用。

公職王歷屆試題 (99 地方政府特考)

(一)使用測驗的必要性:美國心理學會在 1966 年出版的「教育和心理測驗標準」,將心理測驗依其複雜程度分為三級:

1. A 級:用法簡單,使用者可依指導說明手冊之說明加以實施、評分和解釋,一般的教育成就和職業技能測驗多屬此類。具有專業資格之心理師、職能治療師、語言治療師、特教老師或具該測驗研習證書者。
2. B 級:要求使用者條件較嚴謹,使用者需具有測驗編製和使用的技術性知識,且輔以其他相關課程如統計、個別差異、心理學和輔導學等,性向測驗和興趣量表屬此類。社會服務及心理諮商機構之社工師或大學心理、輔導學系等相關科系畢業且修過心理測驗與教育統計,或具該測驗研習證書者。
3. C 級:最複雜要求最嚴謹的測驗,使用者除具有心理測驗及上述各種有關學科之知識外,尚須有專門使用這類測驗設計的課程和實習訓練,以培養使用者的測驗能力和臨床技巧,個人智力測驗、人格量表、投射技術等屬此類。幼稚園教師可在具有特教或測驗專業人士督導之下使用。

(二)選用適當的測驗(王振世、何秀珠、曾文志、彭文松,民 98)

1. 界定測驗需求:當選擇已經出版測驗首要工作,必須將實施測驗目的與測驗結果訊息類型清楚界定,如學校為選用適當的測驗,先確認教學目標、學習內涵與測驗用途。
2. 縮小選擇範圍:測驗使用者必須考量學校是否以實施相關測驗,如以實施正式課堂測驗或評量,則只實施一般教育發展測驗;同時必須考量測驗公平性。
3. 尋找適當測驗:測驗使用者從出版社測驗目錄及最新版「心理測量年鑑」與「測驗評論」,編列出適用測驗清單,淘汰不佳或技術缺陷測驗。
4. 取得樣本題組:測驗出版社會提供較低成本樣本題組,包括指導說明手冊、題本與計分解釋。
5. 檢視測驗材料:測驗使用者必須檢視測驗用途、實施與解釋測驗者的要求資格、測驗信度與效度證據、施測與計分說明、常模與解釋測驗分數依據,最好將每一試題重新檢視其內容。
6. 使用測驗評鑑表:測驗使用者運用測驗評鑑表,蒐集特定測驗訊息,評鑑其一般特徵(評論者意見、優缺點論述、區域使用建議)、技術特徵(信度、效度、常模...)、實用性特徵。

(三)加強測驗發行者與持有者責任:

1. 控制心理測驗使用,一方面防止測驗濫用,一方面基於測驗安全考量;為保持心理測驗價值,測驗發行者及測驗持有者宜對這些工具作安全適當保護。
2. 測驗資料保密性:受試者測驗資料保密性應受到適當保障,其結果只能給有需要人員知道,不可向其他人員透露,同時亦應控制測驗資料蒐集、保存與應用,以免造成受測者不利影響。

【本文參考:1.王振世、何秀珠、曾文志、彭文松(民 98)譯。教育測驗與評量。台北:雙葉書局。2.葛樹人(民 90)。心理測驗學。臺北:桂冠。】

三、有一教育資料庫蒐集了某地區從小一到大四學生的體重資料,其中小一、國一及大一的體重的平均數及標準差如下表:

年級別	平均數	標準差
小一	20.29	2.37
國一	50.78	5.52
大一	61.54	6.25

(一)依上表資料,試就年級別的體重的個別差異由大至小排序。

(二)試舉證支持你排序的理由。

【擬答】:

公職王歷屆試題 (99 地方政府特考)

$$\text{(-)小一 } C.V = \frac{S_x}{X} \times 100\% = \frac{2.37}{20.29} \times 100 = 11.68$$

$$\text{國一 } C.V = \frac{S_x}{X} \times 100\% = \frac{5.52}{50.78} \times 100 = 10.87$$

$$\text{大一 } C.V = \frac{S_x}{X} \times 100\% = \frac{6.25}{61.57} \times 100 = 10.15$$

∴年級別的體重的個別差異由大至小排序：小一>大一>國一。

(二)支持排序的理由：

1. 變異係數(Coefficient of Variation: C.V.)二個或二個以上群體，單位不同欲比較其變異程度，或單位相同但平均數差距太大時，不能以標準差直接比較，因此需以不具單位比例，才可測出其間差異程度。(張子傑，民 86)
2. 適用於比較同一群受試者在不同條件下接受同一測驗時變異情形，或比較兩個不同群體的平均數差別不大時其標準差的不同，不能以標準差直接比較，因此需以不具單位比例，才可測出其間差異程度。
3. C.V. 是相對變異量數中最主要的一個。C.V. 只是描述統計方法，不能用於推論。在推論統計中，欲比較兩標準差是否相同，必須使用變異數差異顯著性考驗。

四、有一企業請專家為其員工講授兩天 16 小時的「資訊安全」課程，且規定要課後評量，評量成績將做為年度考績的參考。業主主觀上認為上課坐前排者較認真，成績應比較好；坐後排者則反之。課後評量後，業主刻意把前兩排及後兩排的成績分開處理，然後請其助理群比較前、後排的得分是否有顯著差異。但其助理群對統計分析的方法有不同的看法。A 助理認為以單因子變異數分析處理會較有深度，而 B 助理則認為用 t 檢定比較簡單，也能達到分析的效果。兩人根據同樣的資料運用電腦軟體分析出來的結果如表 A 及表 B。

Group Statistics

座位	N	Mean	Std. Deviation	Std. Error Mean
前兩排	16	70.25	.931	.233
後兩排	17	69.53	1.663	.403

表 A ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.280	1	4.280	2.318	.138
Within Groups	57.235	31	1.846		
Total	61.515	32			

表 B Independent Ssmples Test

	Levene' s Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
分	Equal	9.040	.005	1.523	31	.138	.721	.473	-.245	1.686
數	variances Assumed									
	Equal Variances not assumed			1.548	25.425	.134	.721	.466	-.238	1.679

公職王歷屆試題 (99 地方政府特考)

試根據前述及表 A、表 B 回答下列問題：

(一)檢定此問題的假設如何陳述？

(二)表 A 及表 B 雖是不同統計方法的結果，但兩者有何種關係存在？

(三)就統計分析的觀點看，你認為那一位助理的方法較妥適，並說明你的理由。

【擬答】：

(一)研究假設：上課坐前兩排的課後評量成績分數應顯著高於上課坐後兩排。

統計假設： $H_0: \mu_1 \leq \mu_2$

$H_1: \mu_1 > \mu_2$

(二)研究上習慣使用 t 考驗解決二組平均數差異比較（二個實驗組的實驗設計），而使用變異數分析處理三組或三組以上的實驗設計，而 F 考驗法亦可處理二組的實驗設計。兩者關係存在

1. $K=2$ 時， $F=t^2$ 。

2. F 考驗（單因子變異數分析）與 t 考驗相同：（余民寧，民 95）

(1) t 考驗用來處理二組平均數之差異比較問題

(2) 變異數分析一般都用來處理三組或三組以上的實驗設計比較各組平均數差異。亦可處理二組平均數之差異比較問題

(3) 兩者均適合小樣本的統計考驗

(4) 二者均為母數統計學，均須符合常態分配、變異數同質性、獨立性

(三) B 助理的方法較妥適。

1. 分析重點放在強調「二個平均數差異是標準誤幾倍」，倍數愈大表二平均數差異愈大，不論是否達顯著水準無須進行事後比較與事前比較，此方法較為簡易進行。

2. B 助理方法首先必須考量兩組獨立母體變異數同質性假設是否違反，可以進行 Levene's 考驗，以變異數分析（F 檢定）為概念，計算兩個樣本變異數比值；F-test 接受

$\sigma_1^2 \neq \sigma_2^2$ ，電腦報表必須看「不假設變異數相等」，即 $t=1.548$ ，P 值 $.134 > .05$ ，落入接受區，亦即無充分證據支持上課坐前兩排的課後評量成績分數應高於上課坐後兩排。A 助理仍須進行 Levene's 考驗，會產生兩張報表，如違反變異數同質性時並無法使用單因子變異數分析。