

# 97 年公務人員高等考試三級考試

類科：教育行政

科目：教育測驗與統計

一、請解釋下列有關測驗的名詞：

- (一)典型表現測驗(typical performance tests)
- (二)實徵效度(empirical validity)
- (三)穩定係數(coefficient of stability)
- (四)試題作答理論(item response theory, IRT)
- (五)適性測驗(adaptive tests)

**【擬答】**

- (一)典型表現測驗(typical performance tests)

在一正常情境中，受試者表現出的典型行為，此一測驗分數無好壞優劣之分。受試者實施測驗時，有時會有說謊或偽飾作答，實際上施行困難度較高。如人格測驗、興趣測驗、態度測驗與適應測驗均屬於典型表現測驗。

- (二)實徵效度(empirical validity)

實證效度為測驗效度建立方式之一，經由另一客觀標準實際驗證後建立，亦即使用該測驗所採用效標，新編測驗實施後得一組分數，效標是此同一群受試者在另一效標測驗分數，求此兩組分數的相關係數即表實徵效度。凡是根據效標建立效度的測驗都可以規類為實徵效度，例如同時效度與預測效度。

- (三)穩定係數(coefficient of stability)

指同一份測驗在不同時間針對相同學生前後重複測量二次，根據二次分數所求得其相關係數，即稱再測信度係數，又稱穩定係數(Coefficient of Stability)。再測信度基本假設測驗所測量潛在特質，在短時間內不隨時間消逝而變化。其誤差來源——來自不同時間下測量所造成之誤差。動作技能測驗、人格測驗與感覺辨別測驗，在短時間內不易變動，適合採用再測信度。

- (四)試題作答理論(item response theory, IRT)

試題作答理論(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增函數描述(即項目特性函數 Item Characteristic Curve, ICC)；ICC 曲線表示，當能力水準  $\theta$  增加時，其反應出回答正確的機率也將增加，又稱試題反應理論。

- (五)適性測驗(adaptive tests)

電腦適性測驗(Computerized Adaptive Testing, 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」理想境界，簡稱「合身測驗」或「適性測驗」。

二、一位教育研究者擬探討學生對學校滿意度與學習成就的關係，於是從某地區之國民中學隨機抽取 20 名學生為樣本，再對這些學生施以學校滿意度問卷，並搜集其學業成績，資料如下(x：滿意度；y：學業成績)：

x：10, 09, 15, 18, 06, 09, 17, 16, 15, 12, 18, 15, 10, 15, 11, 07, 08, 12, 09, 15

y：46, 65, 69, 70, 42, 50, 90, 95, 74, 82, 97, 79, 68, 78, 67, 60, 71, 70, 79, 88

從這些成對的資料可獲得上述兩個變項的描述統計(如表 1)及直線迴歸分析的部分資料(如表 2)：

表 1:

變項	N	Mean	Std. Deviation
滿意度	20	12.35	3.760
學業成績	20	72.00	15.016

表 2:

模式	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(常數)	36.176	8.336		4.339	.000
滿意度	2.901	.647	.726	4.483	.000

依變項：學業成績

請根據以上資料回答下列問題：

(一)寫出以原始資料表示之從滿意度預測學業成績之直線迴歸方程式。

(二)寫出以標準分數表示之從滿意度預測學業成績之直線迴歸方程式。何以表 2 中之標準化係數中缺少常數項？

(三)若設顯著水準(level of significance)  $\alpha$  為 0.05，則(一)與(二)之直線迴歸方程式是否有統計意義？請加以說明之。

(四)若知道某國中生對學校滿意度的分數剛好是上列 20 名學生的平均數，則理論上該生之學業成績為幾分？

(五)若有一國中生對學校滿意度為 13 分，則理論上其學業成績為幾分？

【擬答】

(一)原始分數迴歸方程式  $\hat{Y} = 36.176 + 2.901X$

(二)1. 標準分數迴歸方程式  $\hat{Z}_Y = \beta Z_X = .726 Z_X$

2. 因為預測變項 X 變項與效標變項 Y 變項均化為標準分數，此時截距 a 沒有了，連斜率 b 變成為  $\beta$ ，因此缺少常數項。

(三)1. 常數  $\alpha$  考驗：

$$(1) H_0: \alpha = 0$$

$$H_1: \alpha \neq 0$$

$$(2) t = \frac{a -}{\sqrt{V a}} = 4.339$$

(3) Sig = 0.000 <  $\alpha = .05$  表落入拒絕區，接受  $H_1$  拒絕  $H_0$ ，亦即支持母體迴歸預測線截距  $\alpha \neq 0$ ，惟犯型 I 錯誤 5%。

2. 迴歸係數考驗：

$$(1) H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

$$(2) t = \frac{b -}{\sqrt{V b}} = 4.483$$

(3) P = 0.004 <  $\alpha = 0.05$  表落入拒絕區，拒絕  $H_0$ ，亦即支持母體迴歸預測線斜率  $\beta \neq 0$ ，惟犯型 I 錯誤 5%。

3. 統計分析結果：

以在學成績預測基本學力測驗為一簡單迴歸分析；Pearson 相關係數，Beta 係數皆為 0.726 達顯著水準。係數估計結果指出，在學成績能夠有效預測基本學力測驗，Beta 係數達 0.726 ( $t = 4.483$ ,  $P = 0.000$ )，表示在學成績愈高，基本學力測驗愈好，亦即整體迴歸方程式有顯著預測意義。

$$(四) 國中生 X = \bar{X} = 72 \text{ 代入 } \frac{\hat{Y} - \bar{Y}}{S_Y} = r \cdot \frac{X - \bar{X}}{S_X} \therefore \frac{\hat{Y} - 72}{15.016} = r \cdot \frac{72 - 72}{12.35} \therefore \hat{Y} = 72$$

$$(五) 國中生 X = 13 \text{ 代入 } \hat{Y} = 36.176 + 2.901X = 36.176 + 2.901 \times 13 = 73.889$$

三、下圖係一位統計學老師某次隨堂考試其學生得分的資料所構成的莖葉圖(Stem-and-LeafPlot)。請根據此莖葉圖回答下列問題：

Stem & Leaf

4. 23  
 4. 6  
 5. 00  
 5. 56  
 6. 023  
 6. 5578899  
 7. 0000111234  
 7. 5888999  
 8. 2244  
 8. 566788  
 9. 013  
 9. 578

Stem width : 10

- (一)以莖葉圖展示資料有何優點？
- (二)上圖係由幾名學生的分數所構成？
- (三)此次隨堂考試之最高分與最低分分別為幾分？
- (四)此次隨堂考試分數之眾數(mode)與中數(median)分別為幾分？
- (五)依上圖判斷，此次隨堂考試分數之平均數大於眾數或小於眾數？

【擬答】

- (一)莖葉圖優點包括 1. 圖形兼具數字和圖形的優點； 2. 保留原始資料 3. 莖葉圖橫著看與直方圖功能類似。 4. 可討論數值中是否出現極端值。但如原始數據繁多，則難以計算。
- (二) 50 名。
- (三)最高分=98；最低分=42。
- (四)眾數=70 分；中位數=72.5 (中位數(項次) =  $\frac{50}{2} + \frac{1}{2} = 25.5$  項，介於 72 與 73 之間)
- (五)平均數大於眾數。(平均數=73.44 大於眾數=70；同學可由圖形看出分數比眾數高的分數次數較多或土法煉鋼總分除以 50)

四、請回答下列有關測驗效度的問題：

- (一)試說明效標關聯效度(criterion-related validity)的意義。
- (二)試舉兩種最常用於表示效標關聯效度的類型，並說明其意義。
- (三)如果有三個測量相同特質(如性向)的測驗(測驗 A、測驗 B 及測驗 C)的效度資料如下：

效度類型	效度係數		
	測驗 A	測驗 B	測驗 C
內容效度 (Content validity)	.90	.90	.85
同時效度 (Concurrent validity)	.90	.80	.85
預測效度 (Predictive validity) (相隔一個月)	.50	.65	.60
預測效度 (Predictive validity) (相隔六個月)	.45	.15	.55

上表中那一類型的效度及其相對應的效度係數不合理？請說明為何不合理？

- (四)若刪除妳/你認為不合理的其中一種效度後，再根據以上資料，在三個測驗的其它條件都相同的情況下，如要用來預測某種行為，請將三個測驗的可接受程度由優至劣排序(如 ABC、ACB 或 CBA 等)，並說明妳/你的排序理由。

【擬答】

(一)效標關聯效度是用經驗性(實證分析)方法,研究測驗分數與外在效標關係,探討效度類推性(某個情境結論能夠推論或應用其他情境證據)又稱實證效度、經驗效度或統計效度;測驗成績對目前或未來某一行為表現預測力高低。效標關聯效度分為同時效度與預測效度二種,主要的區別方式在測驗使用目的不同與外在效標取得時間的不同。(郭生玉,民82) 2.測驗結果分數與外在效標相關愈高,表示效標效度愈高,亦即測驗分數能有效預測外在效標,亦即選取效標時最重要考量為預測力。

(二)類型區分:效標關聯效度分為同時效度與預測效度二種,主要的區別方式在測驗使用目的不同與外在效標取得時間的不同。

1.同時效度(concurrent Validity):

(1)指測驗分數與外在效標取得在同一時間內連續完成,計算此二種資料的相關係數即為同時效度。目的在使用測驗分數估計個人在效標方面的實際表現。

(2)例如:新編一國中數學成就測驗實施於一群國中生,以他們在校數學月考成績為效標求其相關,如同時效度適當,則可使用測驗分數瞭解學生的學習表現。

(3)大部分測驗指導手冊中效度,均為同時效度。

2.預測效度(predictive Validity):

(1)指測驗分數與外在效標取得,在相隔一段時間,測驗分數取得在前,而外在效標取得於實施測驗一段時間後,計算此二種資料的相關係數即為預測效度。目的在使用測驗分數預測個人在效標方面的未來表現。

(2)預測效度對於人員選擇與分類測驗最為合適,如學校、工商企業界進行人員甄選與安置。如性向測驗、智力測驗與成就測驗都需採用預測效度,因為三者皆被用來預測未來學習的成就或工作的成績。

(3)大部分測驗手冊上效度,都屬於同時效度,主要原因在於同時效度之效標資料容易取得。

(三)1.應去除內容效度。

2.內容效度是指測驗內容與教材內容、教學目標間,其代表性與取樣適切性。內容效度無數量表示方法(亦即無須進行統計分析),其採用邏輯分析方式,一般都聘請課程專家或審題小組,依據雙向細目表或測驗明細表(Table of Specification)來進行,逐題判斷是否合乎教學目標與教材內。因此不應該有數字出現。

(四)1.三個測驗可接受程度由優至劣排序為CAB。

2.C測驗其同時效度.85與A測驗.9均屬於良好效標關聯效度,而C測驗預測效度為.6與.55比A測驗均,因此C測驗最佳,而B測驗的預測效度經相隔6個月時間後只有.15(極容易受時間影響)因此可接受程度最差。