

## 97 年特種考試地方政府公務人員考試試題

等 別：四等教試

類 科：教育行政

科 目：教育測驗與統計概要

一、某研究者想探討男性與女性的數學能力是否有差異。請回答下列問題：

(一)本研究的虛無假設及對立假設分別為何？

(二)說明本研究中的第一類型錯誤為何？

(三)說明本研究中的第二類型錯誤為何？

【擬答】

(一) $H_0: M_1 = M_2$

$H_1: M_1 \neq M_2$

(二)1. 型 I 錯誤(Type I Error)或第一類型錯誤：當虛無假設 $H_0$ 為真，但依據統計考驗卻拒絕 $H_0$ ，認為 $H_1$ 是對的，稱為型 I 錯誤。其中 $\alpha$ 風險係指犯型 I 錯誤之機率大小，在考驗中又稱顯著水準的風險。

$\alpha = P(\text{拒絕 } H_0 | H_0 \text{ 為真})$

2. 男性與女性的數學能力無差異是真的，卻因統計考驗結果，認為男性與女性的數學能力有差異。

(三)1. 型 II 錯誤(Type II Error)或第二類型錯誤：當虛無假設 $H_0$ 為假，卻因統計考驗結果接受了 $H_0$ ，認為 $H_0$ 是對的，稱為型 II 錯誤。其中 $\beta$ 風險是指犯型 II 錯誤機率大小。

$\beta = P(\text{接受 } H_0 | H_0 \text{ 為假})$

2. 男性與女性的數學能力無差異是假的，卻因統計考驗結果，認為男性與女性的數學能力無差異。

二、卡方統計法適於用來處理人數、次數等間斷變項類料，所以在調查研究法中最常用到。雖然

只有一個定義公式 $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$ ，但其適用情境卻有許多種，請列出其中二種用情境。

【擬答】

(一)適用情境—適合度考驗

教育領域上常使用比較實驗觀察次數與理論之間差異情況，使用「適合度考驗」。只根據單一自變項成分成幾個不同類別或水準(Levels)，稱為單因子分類(One-way Classification)，考驗單因子分類資料的實際觀察次數( $f_o$ )與理論次數( $f_e$ )是否相符，或資料分數分配是否呈現常態分配，使可使用適合度考驗。

(二)適用情境—百分比同質性考驗

在心理與教育研究領域中，將蒐集資料進行 $\chi^2$ 考驗，安排在 I 個橫列和 J 個縱行的方格形狀表格，稱為「交叉表」或「列聯表」(Contingency Table)。細格中資料是次數與人數，皆屬於類別變項。考驗研究對象中 J 組受測者對 I 個反應中，選擇某一反應變項的百分比是否相同。其中 J 組(類)為研究者操控的「設計變項」，I 個反應項目，稱為「反應變項」為研究者所要瞭解或探討的依變項。又稱齊一性考驗。

三、請舉出三個最常見的集中量數，並說明適用時機。

【擬答】

(一)算術平均數

1. 意義

算術平均數(Arithmetic Mean)乃將所有觀察值的數值資料的總和除以總觀察值的個數，由於其計算簡單方便且易於瞭解，是集中量數中最常使用的量數，因此日後統計學上稱呼平

## 公職王歷屆試題 (97 地方政府特考)

均數一般都為算術平均數。簡稱平均數，常  $M$ ， $\mu$ ， $\bar{X}$  (唸 x bar) 表示。

### 2. 算術平均數適用時機：(謝廣全，民 82)

- (1) 用於等距變項及比率變項，推論統計時可進行推論。
- (2) 用於比較兩個團體之成就水準。
- (3) 作其他較複雜統計分析之用。
- (4) 作為測驗常模基準
- (5) 最為穩定集中量數。

### (二) 中位數

#### 1. 意義

中數(Median)又稱二分位數或中位數(Mid-score)，係指順序數列或次數分配中心項之數值，一般皆以  $M_d$  表示。將資料排序(由小到大)，最中間位即表中位數，代表此一團體在此變項的集中情形。中位數等於第二「四分位數」、第五「百分位數」、第五十個「百分位數」。

#### 2. 適用時機

- (1) 中位數為資料最中間數值，不易受極端值影響，感應較不靈敏。因此當一組資料中出現極端值時，使用中位數表示集中量數，比算術平均數更恰當。
- (2)  $\sum |x_i - M_d| \leq \sum |x_i - A|$ 。即表  $\sum |x_i - M_d|$  為  $\sum |x_i - A|$  中最小的。 $A \in R$ 。
- (3) 中位數適用次序變項、偏態分配與資料中出現極端值。
- (4) 可做無母數統計推論，簡單明瞭容易計算。

### (三) 眾數

1. 意義：眾數(Mode)亦為集中量數之一，指連續數列資料中出現次數最多的數值或最多人所得到的分數。一般都以  $M_0$  表示。眾數並非唯一，也可不存在，通常使用於名義變項。

#### 2. 適用時機：

- (1) 適用於有極端值。
- (2) 適用於偏態資料。
- (3) 適用質的資料。眾數是唯一在質的資料中，可求得集中量數。
- (4) 分組次數表有不確定之組距，仍能求得眾數。
- (5) 近似眾數計算，甚為簡單，易於瞭解。

四、在測驗的編製及研究的進行上，我們常會討論到效度的議題，但在不同情形下，效度的意涵可能不同，請解釋下列各項與效度有關的概念，並簡要說明如何提高以下各類「效度」：(一) 構念效度(二) 內在及外在效度(三) 聚斂與區辨效度(四) 增加效度(incremental validity)(五) 交叉驗證(cross-validation)。

### 【擬答】

#### (一) 建構效度：

1. 指測驗能測量理論的概念或特質的程度，換言之，指測驗分數能夠依據某種心理學的理論構念加以解釋的程度，凡用心理學的概念來分析測驗分數的意義皆為建構效度。(郭生玉，民 74，P. 103)
2. 根據社會學與心理學所建構出來描述某種特質的理論，並對測驗分數所做的分析與解釋。(余民寧，民 84)

#### (二) 內在效度與外在效度

1. 內在效度(Internal validity)指研究者建立自變項與依變項的因果關係時，結論中可以完全指出依變項變化是來自自變項變化，而不是其他干擾變項影響。郭生玉(民 88)指出「內在效度」為實驗者所操縱實驗變項對依變項所造成影響真正程度，亦即實驗處理是否確實造成有意義差異。而內在效度高低，則視研究者是否控制無關變項程度，控制愈好，則實驗差異愈能解釋是由實驗處理造成，即內在效度高，亦為一致性高。
2. 外在效度(External validity)指實驗結果的概括性(Generalizability)和代表性

## 公職王歷屆試題 (97 地方政府特考)

(Representativeness)，亦即研究結果是否可以推論到其他研究母體或研究情境，研究結果愈具普遍的應用性，其外在效度愈高。學者 Bracht 與 Glass 提出二種外在效度：1. 母群體效度(Population validity)；2. 生態效度(Ecological validity)。

### (三) 聚斂效度與區別效度

1. 聚斂效度(輻合性效度)(Convergent Validity)：一份測驗分數必須和相同構念其他測量的分數間存在高相關。如美勞性向測驗的分數與在校美勞科目成績有高相關，即為聚斂效度。
2. 區別效度(辨別效度)(Discrimination Validity)：一份測驗必須和不同構念或潛在特質其他測量的分數間存在低相關。如美勞性向測驗分數與在校閱讀理解測驗成績有低相關甚至零相關，即為區別效度。

### (四) 增加效度(Incremental Validity)：

當公司使用測驗甄選員工時，使用增加效度代表測驗使用正確決斷所增加效度；亦即使用測驗做篩選比未使用測驗使篩選所增加效度。增加效果 = 命中率 - 基本率。

### (五) 交叉驗證(Cross-Validation)：

考驗交叉效度需要兩個不同的獨立樣本，及對考驗樣本所訂下的一些效標的測量。這種考驗要回答的問題是，如果在 A 樣本中發現某一些測題得分之組合與效標的相關最大，那麼，在 B 樣本中是否這種關係也同樣會出現。同樣地，在 B 樣本中最有效度的測題組合，在 A 樣本中是否也會出現。請注意這個考驗程序的關鍵在於，首先在 A 樣本中找出與效標相關最大的測題組合，然後再獨立地用 B 樣本去考驗這個組合是否同樣地與效標有高相關。

【本文摘自：陳淑娟、楊中芳譯校(J.P. Robinson、P.R. Shaver & L.S. Wrightsman 著)(民95)。性格與社會心理測量總覽(上)。臺北：遠流出版公司】

五、(一)對客觀式測驗進行試題量化分析，經常會分析題目的那些特性？請說明各特性主要指標的意義與計算方式。

(二)常模參照測驗以及效標參照測驗在選擇試題時，如何運用這些試題指標？

#### 【擬答】

(一) 1. 分析題目特性：量化目的主要在分析每題試題所具備三大統計特徵—難度

(Difficulty)、鑑別度(Discrimination)、誘答力(Distraction)。所有心理測驗中的試題必須經質與量二方面分析，才能顯現其特性與有效程度高低，據此修改或刪除試題，則試題才是一套可靠有效的測驗。

2. 特性主要指標與計算方式

(1) 難度分析

- ① 答對百分比分析：計算全體受試者答對每個試題的人數占全體總人數的百分比值，這個百分比值稱為「難度指標(數)」(Item Difficulty Index)；通過百分比多代表這個題目簡單，通過百分比少代表這個題目較難。

$$\text{公式：} P = \frac{R}{N} \quad , \quad P = \frac{P_H + P_L}{2}$$

- ② 等距量尺分析：美國教育測驗服務社(Educational Testing Service，簡稱 ETS)創用，以 13 為平均數，4 為標準差，以  $\Delta$  (Delta) 代表試題難度指標。下限為 1，上限為 25。

$$\text{公式：} \Delta = 13 + 4Z / [\Delta(\text{Delta}) \text{ 難度指標}]$$

(2) 鑑別力分析：鑑別力分析的目的，主要在瞭解試題是否具備區別學生能力高低的作用。某個試題鑑別度愈高，表示能明確分辨學生答對與答錯功能很強，亦即高分組會傾向答對，低分組會傾向答錯，試題具備此種辨別作用，即為「試題鑑別度」。一個優良的測驗試題，具有較高鑑別度，表示其區辨力愈高。其方法分為三類：

- ① 內部一致性(Internal Consistency)：探討個別試題得分與整份測驗總分間是否合乎一致性，目的在使測驗的內部一致性變得最大。教師自編測驗較難尋找合適的外在效標，此種分析可以改進內容效度與建構效度，但無法增加效標關聯效度，又稱諧度分

析。

$$\text{公式： } D = P_H - P_L$$

②外在效度分析：

- ①分析學生在試題上的反應與在效標上表現之間的關係，目的在試題反應與測驗的外在效度變得最大，用來增強效標關聯效度。
- ②根據學生在某題試題反應答對或答錯與測驗總分相關係數，此一相關係數即用來表示試題鑑別度指標；點二系列相關係數值愈高，代表試題的作用與測驗總分的作用間愈一致，其鑑別功能愈強。即一個「二分變項」(試題分成答對與答錯)與一個「連續變項」(外在效標)。其常用方法：點二系列相關、 $\phi$ 相關、二系列相關法。

(二) 1. 常模參照測驗如何運用試題指標：

(1) 難度分析

- ①  $P = 0.5$  (難度適中)，分配型態—常態分配。
  - ②  $P = 0.3 \leq P \leq 0.7$  (難度適當)。
  - ③  $P > 0.7$  (試題較易)，分配型態—負偏分配。
  - ④  $P < 0.3$  (試題較難)，分配型態—正偏分配。
- $P$  值愈大，難度愈低； $P$  值愈小，難度愈高； $P$  值以 0.5 最恰當 (難易適中)。

(2) 鑑別度分析：

- ①  $D = 0$  (無鑑別度)：表試題太簡單，高分組與低分組學生全部答對；表試題太困難，高分組與低分組學生全部答錯。
- ②  $D = +1$  (教學成功)：高分組學生全部答對、低分組全部答錯。
- ③  $D = -1$  (教學失敗)：高分組學生全部答錯、低分組全部答對。

鑑別度指標值愈高，表示試題鑑別度愈大；鑑別度指標值愈低，表示試題鑑別度愈小。

(3) 選項誘答力分析：客觀測驗進行選項誘答力分析目的在進一步提供試題分析的指標，作為判斷測驗試題編製優劣與診斷學生反應組型的依據。分析選擇題是否具有誘答功能，只要分析高低組學生在每個試題選答的次數分配加以判斷即可。

- ① 每一錯誤選項至少有一低分組學生選擇。
- ② 選擇錯誤選項的低分組學生人數應比高分組學生還多。

2. 標準參照測驗如何運用試題指標：

(1) 鑑別度分析：標準參照測驗主要在測驗受試者的學習成果是否與教學目標相符，有三種常採用方式：①分析受試者在教學前後的測量比較；②分析比較接受教學與未接受教學二組的測驗結果；③比較達到精熟粗或非精熟組二組測量結果。

(2) 難度分析：標準參照測驗是希望大部分學生均能精熟教材，多數題目的難度指標均很高 (通過百分比大)。理論上標準參照測驗難度，應和事先預定通過標準相同，如 85% 的分數為精熟標準時，最適的難度指數應在 0.85 左右。