99 年公務人員普通考試試題

類 科：教育行政
科 目：教育測驗與統計概要

一、解釋並比較下列各對名詞：

㈠性向測驗(aptitude test)與成就測驗(achievement test)
㈡常模參照測驗(norm-referenced test)與效標參照測驗(criterion-referenced test)
㈢難度(difficulty)與鑑別度(discrimination)
㈣比率智商(ratio IQ)與離差智商(deviation IQ)
㈤再測信度(test-retest reliability)與複本信度(alternate form reliability)

【擬答】：

㈠1.性向測驗（Aptitude Test）：性向測驗主要在測量個人學習知識與技能的潛在能力，與預測個體未來發展的可能性。通常包括普通性向（學習一般事務所共同需要的能力）與特殊性向（學習音樂、美術、機械、科學等特殊才能所需的能力）兩種。性向測驗與智力測驗皆為最大表現測驗。

2.成就測驗（Achievement Test）：成就測驗是在測量經學校教育或訓練所獲得的實際能力或表現行為的測驗。成就測驗多半以紙筆方式施行，又稱紙筆測驗。通常包括綜合成就測驗、特殊成就測驗和診斷性測驗三種。

3.兩者均為認知測驗，性向測驗是先天賦予能力，成就測驗則為後天訓練學習能力。

㈡1.常模參照測驗：指測驗結果分數並無意義，必須根據分數在團體中位置而加以解釋，採用相對性標準比較，一般以平均水準為參照點，其目的在競爭區辨學生程度高低。如大學聯考、國家高普考試、標準化成就測驗、智力測驗、性向測驗等。

2.標準參照測驗（效標參照測驗）：根據教學前事先所訂定的絕對性標準加以解釋測驗。一般均以學生所學習知識或技能，判定其「及格或不及格」、「精熟或不精熟」、「通過或不通過」，不需參考他人表現來比較。如國小教師自編測驗（平時考、小考、隨堂測驗）、國家技師執照考試、Bloom 提倡「精熟學習」、中醫師檢定考試、汽車駕照考試。

3.兩者是由解釋測驗分數功能來區分。

㈢1.難度：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），使用次序量尺，可以指出題目等級順序或相對難度。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）。

公式：$P=\dfrac{P_H+P_L}{2}$ 或 $P=\dfrac{R}{N}$

2.鑑別度指數：鑑別度分析的目的，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。

公式：二系列相關 或 $D=P_H-P_L$

3.難度是鑑別度的必要條件，鑑別度是難度的充分條件。當鑑別度 0.4 以上、難度為 0.5 時，此份測驗品質較優良。

㈣1.Binet 在 1908 年的比西量表首先使用心理年齡（Mental Age）表示智力測驗的結果。後來學者為使心理年齡更易解釋，將其轉化成比率智商（Ratio Intelligence Quotient），即心理年齡與實足年齡（Chronological Age）比率，再乘以 100。

$$IQ=\frac{MA}{CA}\times 100\%$$

2.離差智商階段：學者魏克斯勒（Wechsler）創用離差智商（Deviation IQ）首先魏氏採用每個年齡階段內全體智力分布的常態分配，再將個人分數在某年齡組中與平均數差距，用標準差來表示。此離差智商常態分配是以標準分數表示，智商平均數 100，標準差 15。

3.離差智商解決斯比量表不合理現象，人類智力發展隨年齡先快後慢，故 MA 與 CA 比值不可能保持固定值不變，否則成年後 MA 停止增加，但 CA 卻不斷成長增加，其結果必造成兒童智力高，但到成年後智商卻變低的不合理現象。

㈤1.再測信度：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。再測信度適合採用動作技能測驗、視動統整測驗與感覺辨別測驗。

2.複本信度：當同一種測驗有兩種以上的複本測驗（Alternate-form Test）（測驗性質類似在內容、試題格式、長度、難度、鑑別度均一致）時，可將此兩種複本分別實施於同一群受試者，再根據此兩份測驗分數計算其相關係數，即得複本信度，又稱等值係數（Coefficient of Equivalence）。複本信度的測量誤差來源，主要是來自試題抽樣產生的誤差。複本信度適用成就測驗。

二、㈠試從信度的意義說明為何教育人員應該重視測驗工具的信度？㈡測驗的測量標準誤(standard error of measurement)與信度的關係為何？㈢某學生應考某一標準化智力測驗得 95 分，該測驗標準差 15 分，信度 0.91，則計算該智力測驗的測量標準誤 SEm=？㈣並計算該生在該測驗的真正分數 95%信賴區間為多少？(附常態分配表)

【擬答】：

㈠信度是指相同受試者在不同時間，使用相同測驗測量（或複本測驗測量多次）或在不同情境下測量，所得結果一致性。藉以反應真實量數程度的一種指標。二次測量結果相當一致，就表示測量分數可靠性高、穩定性高或具有可預測性。危芷芬（民 88）提出信度為受試者在不同場合、採用內容相當，不同題目、不同測試條件下，重複接受測驗所得分數一致性；一份優良教育測驗信度至少要在.80 以上的信度係數才有使用價值。

㈡測量標準誤與信度係數關係：

1.$r_{XX}=1$ 時，$SE_{meas}=0$；$r_{XX}=0$， $SE_{meas}=S_X$（測驗分數的標準差）。r_{XX}與 SE_{meas}二者呈上升、下降曲線關係。

2.信度愈高，測量標準誤愈小。而標準差 S_X與 SE_{meas}會隨信度係數固定時，二者成正向關係。

㈢$SE_{meas}=S_X\sqrt{1-r_{XX}}=15\sqrt{1-0.91}=4.5$

㈣$X-Z_{.975}\cdot SE_{meas}\le t\le X+Z_{.975}\times SE_{meas}$

$95-1.96\times4.5\le t\le95+1.96\times4.5$

$86.18\le t\le103.82$

進行區間估計該生在測驗真正分數介於 86.18 至 103.82 之間可信賴程度為 95% 。惟犯型一錯誤 5% 。

三、解釋並比較下列各對名詞：

㈠連續變數(continuous variable)與間斷變數(discrete variable)

㈡z 分數與 T 分數

㈢描述設計(descriptive statistics)與推論統計(inferential statistics)

㈣第一類型錯誤(Type Ⅰ error)與第二類型錯誤(Type Ⅱ error)

㈤離均差平方和與變異數

【擬答】：

㈠連續變項與間斷變項：

1.連續變項（Continuous Variable）：觀察或測量的統計資料中，任何二數量間均可無窮細分，可以有小數或分數連續不間斷的數據，且每個數值均有意義，連續變項是一個數列中一段距離而非一個點，可使用「介於什麼之間」表示。如身高、時間、體重、分數……等

。

2.間斷變項（Discrete Variable）：某一特定範圍內一些分數點相互分開，且各點之間有一段明確距離，同時此種變項大多使用整數表示；其數值間不可無限細分，其值由點計而得，而非一段距離，不可以有小數或分數。如：性別、骰子點數、家庭孩子數、20 張椅子是一精確數。

3.連續變項適用母數統計學、間斷變項適用無母數統計學；連續變項可以轉化為間斷變項。

㈡1.Z 分數：表示某一測驗分數在團體平均數以上或以下多少個標準差；將原始分數加以轉換成可以互相比較之分數。當二變數欲比較大小，同時考量到不同單位、不同平均水準與不同變異（離散）程度時，就可採用標準分數。係以平均數為參照點，以標準差為單位的相對地位量數。

2.T 分數：由美國測驗專家麥考爾（McCall, 1939）發明，平均數 50、標準差 10；T 分數是 Z 分數經由直線性轉化的標準常態分數，即母體是常態分配。早期是用來瞭解國小學生閱讀測驗分數，蒐集 500 個 12 歲小學生在閱讀測驗的得分，製成 T 量表。T 分數和 Z 分數一樣，使用上較方便，可避免分數有負數與小數。

3.$T=10Z+50$。

㈢1.敘述統計學（Descriptive Statistics）：只探討母體與樣本的個別特性，對研究資料之處理主要在蒐集、整理劃記、描述與表現結果，即將一群資料加以整理、摘要、組織與簡化，使讀者容易明瞭其中意義與傳遞訊息。內容包括集中趨勢、變異趨勢、圖形表示、相對地位量數等。

2.推論統計學（Inferential Statistics）：係探討母體與樣本之間的關係，根據樣本的資料推論未知母體性質（非樣本性質）。當研究者欲瞭解母體某些特性，但母體群過於龐大，無法將母群體全部蒐集，最好辦法為經由隨機（Random）方式，自群體抽取若干具有代表性（Representative）個體為樣本，再依機率原理，經由假設、估計、考驗步驟，使用樣本統計量從事估計與考驗母體的方法。內容包括機率分配、抽樣方法、估計、考驗、卡方分配、相關與迴歸。

㈣1.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 α 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。

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

2.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 β 風險是指犯型Ⅱ錯誤機率大小。

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

3.當其他條件不變情況下，α 與 β 具有互相消長關係，如 α 值增加，則 β 值減少。

㈤1.變異數（Variance）係一群數值與其算術平均數之差異平方和的平均數，稱為變異數，又稱變方或均方變異數，開根號即為標準差。二者為用途最廣的變異量數。當群體中標準差愈小，即表群體中大部分數值集中於平均數附近，則平均數代表性強；相反地，若標準差大則表大部分數值比較分散，平均數代表性較弱。

2.離均差平方和係一群數值與其算術平均數之差異平方和，具有最小平方法特性，亦即

$\sum_{i=1}^{n}(x_i-\bar{x})^2 \le \sum_{i=1}^{n}(x_i-a)^2$ 。

3.離均差平方和與變異數兩者成正比，表示當離均差平方越大則變異數越大。

$$S^2=\frac{\sum(x-\bar{x})^2}{n}=\frac{SS_x}{n}$$

四、假設某校學生的課後補習程度(X 變項)與學業成就(Y 變項)皆為常態分配，自其中隨機抽樣 10 名學生的課後補習與學業成就資料如下，並計算一些統計基本數量如方格內所示，試計算下列各題：

X	Y
8	9
12	10
6	4
8	5
9	6
8	8
10	9
7	7
5	5
5	7

$\Sigma X=78, \Sigma X^2=652$
$\Sigma Y=70, \Sigma Y^2=526$
$\Sigma XY=573$

㈠X 與 Y 之共變數？(5 分)積差相關係數？
㈡Y 變項的變異量不能由 X 變項解釋的變異量為多少百分比？
㈢算出以 X 變項預測 Y 變項的迴歸方程式。(註：須列式才給分)

【擬答】：

㈠1. 共變數 $S_{XY}=\dfrac{\sum(X-\bar{X})(Y-\bar{Y})}{N-1}=\dfrac{\sum XY-\dfrac{\sum X\sum Y}{N}}{N-1}=\dfrac{573-\dfrac{78\times 70}{10}}{10-1}=3$

$$2.\ r=\frac{\sum XY-\dfrac{\sum X\sum Y}{N}}{\sqrt{\sum X^2-\dfrac{(\sum X)^2}{N}}\sqrt{\sum Y^2-\dfrac{(\sum Y)^2}{N}}}=\frac{573-\dfrac{78\times 70}{10}}{\sqrt{652-\dfrac{78^2}{10}}\sqrt{526-\dfrac{70^2}{10}}}=.6815$$

㈡$1-r^2=1-.6815^2=.6835$ ，表示 Y 變項的變異量不能由 X 變項解釋的變異量為 68.35%。

㈢$b_{Y\cdot X}=\dfrac{SS_{XY}}{SS_X}=\dfrac{\Sigma XY-\dfrac{\Sigma X\Sigma Y}{n}}{\Sigma X^2-\dfrac{(\Sigma X)^2}{n}}=\dfrac{573-\dfrac{78\times 70}{10}}{652-\dfrac{(78)^2}{10}}=.6193$

$a_{Y\cdot X}=\bar{Y}-b\times\bar{X}=7-0.6193\times 7.8=2.1695$

$(\bar{Y}=\dfrac{\Sigma Y}{n}=\dfrac{70}{10}=7$，$\bar{X}=\dfrac{\Sigma X}{n}=\dfrac{78}{10}=7.8)$

∴直線迴歸方程式 $\hat{Y}=.6193X+2.1695$