

99 年公務人員高等考試三級考試試題

類 科：教育行政、技職教育行政

科 目：教育測驗與統計

一、請回答下列有關直線迴歸的問題：

(一)在簡單直線迴歸中，一旦標準化後，請問其常數是否存在？請說明之。

(二)在簡單直線迴歸中，若標準化迴歸係數為 β ，原始迴歸係數為 b ，而預測變項與效標變項相關係數為 r ，請問以上三係數或兩係數間有關係存在嗎？請說明之。其中 S_X 為 X 變項的標準差， S_Y 為 Y 變項的標準差。

(三)承上題(二)，請說明該如何了解預測變項對效標變項的解釋力。

【擬答】：

(一)簡單直線迴歸中標準化後，常數因為標準化後已經不具備單位，因此其常數不見了。

$$\hat{Y} = a + bX = (\bar{Y} - b\bar{X}) + bX \rightarrow (\hat{Y} - \bar{Y}) = r \times \frac{S_Y}{S_X} (X - \bar{X}) \therefore \frac{\hat{Y} - \bar{Y}}{S_Y} = r \times \frac{X - \bar{X}}{S_X}$$

$$\therefore \hat{Z}_Y = r \cdot Z_X \text{ [截距 (常數) } a \text{ 不見了]}$$

(二)1. 簡單直線迴歸中 $\beta = r$ 稱為「標準分數迴歸係數」，由於標準化結果 β 係數具有與相關係數相似性質，其值介於 -1 與 1 之間，當絕對值越大時表該變項的預測力越強，而正負號代表 X 與 Y 變項關係方向。2. 迴歸係數 b 表示迴歸方程式預測方向性，如值為正值表示自變項對依變項影響為同方向，反之則相反；迴歸係數帶有單位非標準化統計量，可以反應自變項對依變項影響的數量，但由於單位差異無法進行相對比較。3. $b = r \times \frac{S_Y}{S_X}$ ，亦即 b 與 r 兩者同號且成正比，當 S_X 與 S_Y 兩者不變條件下， b 值越大則 r 值

越大，表示預測變異百分比越大。

(三)使用決定係數 (Coefficient of Determination) 以 r^2 表示，表效標變項 Y 中之總變異百分比有多少可由預測變項 X 來解釋到的變異量百分比，或由 X 變項預測 Y 變項預測正確部分；亦即表示 Y 變項被自變項所削減的誤差百分比 (Proportioned reduction in error; PRE)。因此在總離均差平方和中，迴歸離均差平方和占其多少百分比，便為判斷預測效果大小重要指標。

$$\text{公式} : r^2 = \frac{SS_{\text{reg}}}{SS_t} = \frac{\text{預測正確部分}}{\text{平均數預測Y變項的誤差}}$$

二、實驗一：某研究者想了解三種教學方法對國中生數學科成績的影響，以隨機分派的方式將學生分派到教師中心、學生中心、團體教學等三組教學情境。每組 5 人共 15 人。這些國中生參加一年後，數學科成績如表 1 所示：

表 1：實驗一

教學方法		
教師中心	學生中心	團體教學
4.00	5.00	9.00
3.00	7.00	8.00
5.00	4.00	9.00
7.00	6.00	6.00
6.00	5.00	8.00

實驗二：與實驗一非常相似，唯其在學生分派上有所不同。

某研究者想了解三種教學方法對國中生數學科成績的影響，此實驗為節省人力，故只用五名學生，而每位學生必須經歷三種教學方法(教師中心、學生中心、團體教學)。這些國中生參加一年後，數學科成績如表 2 所示：

表 2：實驗二

學生	教學方法		
	教師中心	學生中心	團體教學
A	4.00	5.00	9.00
B	3.00	7.00	8.00
C	5.00	4.00	9.00
D	7.00	6.00	6.00
E	6.00	5.00	8.00

問題：

- (一)請問實驗一與實驗二各為何種實驗設計？
- (二)分別羅列實驗一與實驗二的研究假設。
- (三)請分別羅列實驗一與實驗二的變異數摘要表。請寫出變異來源、自由度，並以代號方式表示各離均差平方和，以此計算均方和 F 值。

變異來源	離均差平方和	自由度	均方	F
------	--------	-----	----	---

(四)請問實驗一和實驗二何者較容易得到 F 值的顯著，或是兩者並無差異，為什麼？

【擬答】：

- (一)實驗一為完全隨機化設計（受試者間設計）---單因子獨立樣本變異數分析；
實驗二為隨機化區組設計（受試者內設計）---單因子相依樣本變異數分析。
- (二)實驗一的研究假設：國中生不同教學方法對數學科成績有顯著影響。
實驗二的研究假設：國中生重複接受不同教學方法對數學科成績有顯著影響。
- (三)1. 實驗一

教學方法			
教師中心	學生中心	團體教學	
4.00	5.00	9.00	
3.00	7.00	8.00	
5.00	4.00	9.00	
7.00	6.00	6.00	
6.00	5.00	8.00	
$\bar{X} = 5$	$\bar{X} = 5.4$	$\bar{X} = 8$	$\bar{\bar{X}} = 6.1333$

- (1) $H_0: \mu_1 = \mu_2 = \mu_3$
 $H_1: \mu_i: \text{不全等}$

(2)計算：

$$SS_t = \sum \sum X^2 - \frac{(\sum \sum X)^2}{n \cdot k}$$

$$SS_t = (4^2 + 3^2 + \dots + 6^2 + 8^2) - \frac{(4+3+\dots+6+8)^2}{5 \times 3}$$

$$SS_t = 612 - \frac{(92)^2}{15} = 47.7333$$

$$SS_b = n_j \sum (\bar{X}_j - \bar{\bar{X}})^2 = 5[(5 - 6.1333)^2 + (5.4 - 6.1333)^2 + (8 - 6.1333)^2] = 26.5335$$

$$SS_w = SS_t - SS_b = 47.7333 - 26.5335 = 21.1998$$

變異來源	離均差平方和	自由度	均方	F
SS _b	26.5335	2	13.2668	7.5094
SS _w	21.1998	12	1.7667	
SS _t	47.7333	14		

2. 實驗二

學生	教學方法			\bar{X}_i
	教師中心	學生中心	團體教學	
A	4.00	5.00	9.00	6
B	3.00	7.00	8.00	6
C	5.00	4.00	9.00	6
D	7.00	6.00	6.00	6.3333
E	6.00	5.00	8.00	6.3333
\bar{X}_j	5	5.4	8	$\bar{\bar{X}} = 6.1333$

(1) $H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_i$ 不全等

(2) 計算：

$$SS_t = \sum \sum X^2 - \frac{(\sum \sum X)^2}{n \cdot k}$$

$$SS_t = (4^2 + 3^2 + \dots + 6^2 + 8^2) - \frac{(4+3+\dots+6+8)^2}{5 \times 3} = 612 - \frac{(92)^2}{15} = 47.7333$$

$$SS_{b.subjects} = n_j \sum (\bar{X}_j - \bar{\bar{X}})^2$$

$$= 3[(6 - 6.1333)^2 + (6 - 6.1333)^2 + (6 - 6.1333)^2 + (6.3333 - 6.1333)^2 + (6.3333 - 6.1333)^2] = .3999$$

$$SS_{\omega.subjects} = SS_t - SS_{b.subjects} = 47.7333 - .3999 = 47.3334$$

$$SS_{b.treatments} = n_i \sum (\bar{X}_i - \bar{\bar{X}})^2$$

$$= 5[(5 - 6.1333)^2 + (5.4 - 6.1333)^2 + (8 - 6.1333)^2] = 26.5335$$

$$SS_{residual} = SS_{\omega.subjects} - SS_{b.treatments} = 47.3334 - 26.5335 = 20.7999$$

(三) 編製變異數分析表：

變異來源	SS	df	均方	F 值
受試者間 $SS_{b.subjects}$.3999	4		
受試者內 $SS_{\omega.subjects}$	47.3334	10		
處理效果 $SS_{b.treatments}$	26.5335	2	13.2668	$F = \frac{13.2668}{2.5999} = 5.1028$
殘差 $SS_{residual}$	20.7999	8	2.5999	
全體 SS_t	47.7333	14		

(四) 在相依樣本中 (實驗二)，真正誤差為扣除為受試間個別差異造成變異所剩殘差，其數值比獨立樣本的組內誤差 (SS_{ω}) 數值小，因此在進行 F 考驗時，由於分母殘差值較小，相依樣本比獨立樣本 (實驗一) 更易達到顯著水準。

三、請將下列各小題，就選目性質、鑑別指數及難度指數說明試題分析，並決定每一小題是否應修正、刪除或採用？

(一)

組別	選目					鑑別指數	難度指數
	A*	B	C	D	未答		
高分組	16	0	0	4	0	0.10	0.65
低分組	12	0	6	2	0		

公職王歷屆試題 (99 高普考)

*為正確答案

(二)

組別	選 目					鑑別指數	難度指數
	A	B	C*	D	未答		
高分組	7	0	10	3	0	0.29	0.5
低分組	4	7	7	2	0		

*為正確答案

(三)

組別	選 目					鑑別指數	難度指數
	A	B*	C	D	未答		
高分組	4	6	5	5	0	0.07	0.35
低分組	5	5	4	6	0		

*為正確答案

(四)

組別	選 目					鑑別指數	難度指數
	A	B*	C	D	未答		
高分組	0	10	10	0	0	0.35	0.4
低分組	4	6	6	4	0		

*為正確答案

【擬答】：

- (一) 1. 選目性質：B 選項無誘答作用，因為高分組與低分組均無人作答；D 選項高分組選錯人數高於低分組選錯人數表無誘答力（吸引高分組學生作答、可能題目不清、爭議答案、作答粗心）；C 選項低分組有多人作答，但高分組無人選填，此選項有誘答作用。
 2. 鑑別指數：美國測驗專家 Ebel 在 1979 年提出一套評鑑標準，當 D 值在 0.2 以下顯示此題劣、必須加以淘汰。
 3. 難度指數 $P = .65$ 說明此題題目較易。
- (二) 1. 選目性質：A 與 C 選項高分組選錯人數均高於低分組選錯人數表無誘答力；B 選項低分組有多人作答，但高分組無人選填，此選項有誘答作用。
 2. 鑑別指數：美國測驗專家 Ebel 在 1979 年提出一套評鑑標準，當 D 值在 0.2~0.29 顯示此題尚可，通常須修改。
 3. 難度指數 $P = .5$ 說明此題題目難易適中。
- (三) 1. 選目性質：A、B、C 選項高分組選錯人數均低於低分組選錯人數表有誘答力。
 2. 鑑別指數：美國測驗專家 Ebel 在 1979 年提出一套評鑑標準，當 D 值在 0.2 以下顯示此題劣、必須加以淘汰。
 3. 難度指數 $P = .35$ 說明此題題目難度較高題目較難。
- (四) 1. 選目性質：選項 A 與 D 其高分組答錯人數均低於低分組答錯人數顯示具有誘答力，而 C 選項其高分組答錯人數高於低分組答錯人數顯示無誘答力。
 2. 鑑別指數美國測驗專家 Ebel 在 1979 年提出一套評鑑標準，當 $D = .35$ 顯示此題優良，可能須修改。
 3. 難度指數 $P = .4$ 說明此題難度適中稍難。

四、請解釋下列有關的測驗名詞：

- (一) 反應心向(response set)
 (二) 構念效度(construct validity)
 (三) 決策效度(validity of decisions)
 (四) 測量標準誤(standard error of measurement)
 (五) 量尺分數的等化(equating)

【擬答】：

- (一)依照某種習慣性的反應型態，對測驗試題作一致性傾向的反應行為，相同題目以不同方式出現；葛樹人(民88)提出反應心向為受測者有意識或無意識地變更其在測驗反應，而塑造出一種他或她內心中所希望呈現形象，可以採用強迫選擇式的題目減少反應心向作用。例如：猜測、重速度、輕正確、重正確、輕速度均為反應心向類型；反應心向類型包括：猶豫性、隨機反應性、默從性、社會期許。
- (二)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95)學者Loevinger在1957年便提倡建構效度為效度全部內涵，因為預測效度、同時效度及內容效度本質上都在測驗建構前就存在，其性質較不符合科學探究精神，如從科學觀點具有建構效度是完整效度。
- (三)測量工具之效度檢驗，通常可分為兩方面，即測量效度(測量個人的某種屬性)與決策效度。決策效度即預測未來結果並作出個人有關的決策，一般分為人員甄選時決策效度，分析基準率、選擇比率、命中率與增進效度，另一為診斷工具的研究決策效度，結果常以兩種方式呈現：1. 檢驗或檢查的準確度，包括敏感度、特異度，2. 檢驗或檢查在某一群體的預測值，包括陽性預測值(PPV)、陰性預測值(NPV)。
- (四)用以估計受試者測驗之真實分數所在範圍的誤差單位與分數穩定性的信度。在標準情境下，使用相同測驗或複本測驗測量一個受試者許多次，其所得分數的平均數即為個人的真實分數。一般來說，會以受試者的真實分數為中心形成常態分配。而每次測驗實得分數與真實分數的差，稱為測量誤差，這許多次測量誤差分配之標準差即為測驗標準誤，又稱為分數的標準誤差。
- 公式： $SE_{meas} = S_x \sqrt{1 - r_{xx}}$
- (五)國際上大型標準測驗通用的方式，因為具有和群體比較的意義，好像一把用來測「量」個人表現的「尺」，所以稱為「量尺分數」，量尺分數的等化(equating)為將題庫中的每道試題都透過「試題反應理論」(Item Response Theory, 簡稱IRT)的估計程序，將試題難度連結在同一把尺上，所以每道題目的難度值是已知的。著名的TOEFL、GRE、國中基測等電腦化適性測驗也都是採用這種作法。