

# 《教育測驗與統計》

## 試題評析

今年的考題雖然沒有具體的統計計算結果，但是一如預期，教育統計學考的是迴歸分析與變異數分析，這兩種統計方法都是班內同學所熟悉的。特別是第一題迴歸係數與相關係數之間的公式推導關係，以及第二題變異數分析摘要表，在講義中都提供了足夠練習，並充份講解。而第三題的試題分析與第四題的解釋名詞，應也難不倒本班同學。倒是第二題的作答，題目並未要求真正算出最後數值，而是在表格中以符號呈現，是較特別的考法，足見觀念的了解比計算的熟練更加重要！

因此，要在今年超低錄取率的本類組高分上榜，相信在其他科目不失常的情況下，以本科目的難易度，起碼要有85分以上的表現，才較有機會。

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。(5分)
- (二)在簡單直線迴歸中，若標準化迴歸係數為 $\beta$ ，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。(5分)
- (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。(5分)

**答：**

(一)在簡單直線迴歸中，標準化後，原本的迴歸係數將由 $b_1$ 變成 $\beta$ ，而迴歸常數將不復存在，因為常數的標準

化使用的公式 $Z = \frac{b_0 - \bar{b}_0}{S_{b_0}}$ 中，分子的差值為0，因此標準化直線迴歸模式將成為 $Z_{\hat{y}} = \beta Z_x$ ，也就是迴歸常數等於0。

(二)1. 當標準化迴歸係數為 $\beta$ ，原始迴歸係數為 $b_1$ ，兩者之間的關係式須利用兩個變項的標準差，其為

$$\beta = b_1 \frac{S_x}{S_y} \quad (\text{式一})$$

2. 又因為原始迴歸係數 $b_1$ 的計算式為「離均差交乘積和」除以「 $X$ 變項的離均差平方和」，

$$\begin{aligned} b_1 &= \frac{SS_{xy}}{SS_x} \\ &= \frac{cp}{SS_x} && \text{在將分子分母同時除以自由度}(n-1)\text{後，可得} \\ &= \frac{cp/(n-1)}{SS_x/(n-1)} \\ &= \frac{C_{xy}}{S_x^2} && (\text{式二}) \end{aligned}$$

3. 將(式二)代入(式一)，可得

$$\begin{aligned} \beta &= b_1 \frac{S_x}{S_y} \\ &= \frac{C_{xy}}{S_x^2} \frac{S_x}{S_y} \end{aligned}$$

$$= \frac{C_{xy}}{S_x S_y} = r$$

因此，標準化迴歸係數  $\beta$ ，等於原始迴歸係數  $b_1$  乘以兩個變項之標準差的比值，而最終將因算式中的約分，得到相關係數。

- (三)欲瞭解預測變項對效標變項的解釋力，可以由兩個變項的相關係數著手。因為積差相關係數的值域界於-1與1，愈接近正一與負一的相關係數值，表示兩個變項的相關強度愈高。以簡單直線迴歸分析而言，將有正有負的相關係數取平方，可得決定係數  $r^2$  或  $R^2$ ，其表示效標變項的總變異中，可以利用預測變項的數值與迴歸模式加以解釋的百分比。其值愈高，預測變項對效標變項的解釋力也就愈高。

【參考書目】

傅立葉老師，教育測驗與統計講義，第一回第八章P.44-46及上課筆記。

- 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？(5分)
- (二)分別羅列實驗一與實驗二的研究假設。(5分)
- (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。(20分)

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？(5分)

**答：**

(一)實驗一以隨機分派方式將15名學生分至三種不同的教學情境，由於每一種教學情境中的學生只接受該種教學情境，這樣的互為獨立情境，屬於獨立樣本設計，或稱受試者間設計；至於實驗二，因為在節省人力的考量下，5名學生前後分別經歷三種教學方法，並在一年後，每人得出三個數學科成績，所以其實驗設計屬於重複測度(repeated measures)的相依樣本設計，又稱受試者內設計。

(二)實驗一的研究假設為：

- 1.常態母體：樣本取自常態化的母群體；
- 2.變異數同質性：每一個常態樣本具有相同的離散程度；
- 3.變異來源可加性：各種變異來源的變異量需互相獨立，且可以進行累積和加減。

實驗二的研究假設，除了上述的常態母體與變異來源可加性的兩個假設外，因學生參與三種實驗情境的時間歷程，受試者流失與受試者偏見態度的控制，皆為足以影響結論所需特別注意者。

(三)實驗一的變異數分析摘要表如下：

變異來源	離均差平方和	自由度	均方	F
組間變異	$MS_{b.treatment}$	2	$MS_{b.treatment}/2$	$6 MS_{b.treatment} / MS_{w.treatment}$
組內變異	$MS_{w.treatment}$	12	$MS_{w.treatment}/12$	
總和	$MS_{total}$	14		

實驗二的變異數分析摘要表如下：

變異來源	離均差平方和	自由度	均方	F
受試者間變異	$MS_{b.subject}$	4	$MS_{b.subject}/4$	$2 MS_{b.subject} / MS_{error}$
受試者內變異	$MS_{w.subject}$	10	$MS_{w.subject}/10$	$8 MS_{w.subject} / 10 MS_{error}$
處理間變異	$MS_{b.treatment}$	2	$MS_{b.treatment}/2$	$4 MS_{b.treatment} / MS_{error}$
誤差變異	$MS_{error}$	8	$MS_{error}/8$	
總和	$MS_{total}$	14		

(四)以相同的資料(數學科成績)來說，實驗二較易得到顯著的F值，因為重複測度的實驗結果，可以排除受試者間的差異，其可能是對於數學科原本就存在的能力差異或是喜好與否，而將能夠有效釐清介於不同教學情境下，受試學生真正可以在數學科的學習上，所呈現最高的學習效果。

**【參考書目】**

傅立葉老師，教育測驗與統計講義，第二回第十三章P.38-39及上課筆記。

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？(每小題5分，共20分)

(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		

\*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

**答：**

以下分別針對各題，進行難度、鑑別度指數的計算分析，並且進行選目分析；

- (一)第一題的難度屬於中間偏易，但是0.1的鑑別度偏低，因而鑑別受試者的能力不足。選目分析上，相對於選項C，B選項的誘答力不足，因此無人選擇其為正確答案；而選項D因高分組選擇的人數高於低分組，施測者或測驗編製者應修改其文字。
- (二)第二題的難易適中，而0.29的鑑別度屬於可以接受的水準，其足以鑑別受試者能力高低。選目分析上，因為選擇選項A與D作為正確答案的高分組人數多於低分組，施測者或測驗編製者應修改其文字，以提高其選目品質。
- (三)第三題的難度稍高，因而影響其鑑別度，而低的鑑別度可以由高低分組之受試者平均在四個選項「猜答」獲得證據，所以此題並無鑑別受試者能力高低的能力。選目分析上，錯誤的A、C、D選目應進行必要的文字修改，甚至直接抽換，以提高其選目品質。
- (四)第四題0.4的難度接近適中，因而有0.35不錯的鑑別度。選目分析上，也只有選目C因高分組選答的人數高於低分組，違反誘答項的假設，因而需要進行修改。

**【參考書目】**

傅立葉老師，教育測驗與統計講義，第三回第十八章P.26-27及觀念練習題20與模擬試題8。

四、請解釋下列有關的測驗名詞：（每小題6分，共30分）

- (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)

**答：**

- (一)反應心向是指受試者有意或無意地變更其在測驗上的反應，而塑造出他內心所希望呈現的形象，但此形象並不代表真正的自己。反應心向的特徵為，其通常與測驗內容有關，受試者可從測驗內容判斷應如何變更其反應，也常會受到社會期望或受試者之特殊目的影響。其類型有：猶豫型、隨機反應型、默從型與社會期許型。
- (二)構念通常指的是心理特質，其為潛藏在人內在深處的特質，也是一個人外在行為表現的重要原因。構念效度指一測驗能夠正確測量到理論上的構念或特質的程度，又稱建構效度。通常在測量前所未有的、或是以前測的不是很好的事物時，特別注重。因此，在各種測驗中，又以人格測驗的編製特別重視。求得構念效度的方法有：內部一致性分析、多項特質與多項方法分析、因素分析等。
- (三)測驗能夠正確測量目標特質或能力的程度，施測者能夠據以進行教學方法或教材的相關決策，因此，決策效度可以是一個測驗的效標關聯效度的體現。
- (四)通常一位受試者的真正分數為未知，頂多只能以多次測驗結果、實得分數的平均數作為其估計值。而在多次測量後，每次實得分數與真正分數之差值(測量誤差)的分配的標準誤，也就是在一個團體的測驗分數中，全體誤差分數的標準差估計值，稱為測量標準誤(SE<sub>meas</sub>)。其值不僅對於測驗信度的計算相當重要，也常用於進行個別受試者真正分數之信賴區間估計。
- (五)量尺分數的等化，所指為不同量尺分數之間，原本缺乏可以互相比較的基礎，等化後的量尺分數才有具備客觀相對比較的能力。以國中基本學力測驗為例，基本學力測驗每年舉辦兩次，考生可從這兩次測驗中擇優參加登記分發入學，因此必須有一個共同的參照標準，使兩次測驗分數可以互相比較。目前的做法是利用每年參加第一次各科測驗的考生分數為常模，來建立各科答對題數與量尺分數的對照表。同年第二次測驗各科並不再重新計算量尺，而是透過測驗等化的機制，將第二次測驗各科的答對題數對應到依據第一次測驗結果所建立的參照標準上。兩次基本學力題本的難度雖然有些微差異，但是透過共同的參照標準，兩次測驗的分數是可以互相比較的。

**【參考書目】**

傅立葉老師，教育測驗與統計講義，第三回各章及上課講義。