

《教育測驗與統計》

試題評析

今年的考題特別簡單，一定出乎所有考生的意料之外。但也正符合本班上課時所強調，不要一味的鑽研太困難或複雜的考古題。

分析考題型式，除解釋名詞全屬於測驗學(本班教材第三回)外，第二題考的是迴歸分析，第三題考的是統計圖表，也都在本班教材第一回中明示，並曾透過練習題與考古題加以練習解說，第四題考的是效度，雖然其第三小題與第四小題稍需臨場仔細分析，且較有利於有實務經驗者，但是本班同學只要曾在上課時用心聽講，在此科考試應可輕易取得80分以上的好成績。但是，正因為試題太過容易，考75分以下的考生可能上榜不易！

一、請解釋下列有關測驗的名詞：(每小題5分，共25分)

- (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)

答：

- (一)典型表現測驗以主試者的角度，指的是就受試者的喜好或實際行為，希望將受試者作適當歸類的一種測驗。其測量內容較常見的有人格、氣質、興趣、喜好和價值等，基本上，它無所謂「對、錯」或「是、非」之分。主試者希望受試者能坦白誠實的在測驗上確實表現自己，但因為典型表現測驗的分數意義很難說明，常需測驗更多的試題。
- (二)欲了解一個新編測驗預測某一個體行為的有效程度，而研究測驗分數與外在效標間之關聯性所計算而得的相關係數即為效標效度，因為效標的選擇多依研究者的專業經驗判斷，又稱實徵效度。
- (三)在重測法下，使用同一測驗，針對同一群受試者，在不同時間內前後測兩次，求兩次得分間的相關係數(可依資料類型選擇計算積差相關係數或等級相關係數)，稱穩定係數或再測信度(Test-retest reliability)。此種信度係數易受兩次測驗間隔時間長短的影響，不同的測驗種類有其不同的間隔時間要求。
- (四)試題反應理論主要是以個別試題的觀點，來解釋測驗分數的涵義。它認為學生在某一試題上的表現情形，與其背後的某種潛在特質(即能力)之間具有某種關係存在，該關係可以透過一條連續性遞增的數學函數來加以表示和詮釋，這個數學函數便稱作試題特徵曲線(item characteristic curve，簡稱ICC)。
- (五)是一種可以依受試者的程度，由題庫中挑選出最適題目的組合，而測得學生真正的學科能力之測驗，通常是以電腦施測的標準化測驗。在龐大題庫的支持下，因其呈現給每一位受試者的試題難度與順序，是依據個別受試者在測驗一開始，前幾個試題的表現好壞結果來做試題的篩選與決定，因此對每一位考生而言，其受測的試題組合不盡然完全相同。(總複習教材p. 41解釋名詞)

【參考書目】

1. 傅立葉老師《測統教材講義》第三回，與上課筆記。
2. 傅立葉老師《測統總複習教材》，與上課筆記。

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)：(每小題5分，共25分)

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 (未標準化係數)		Standardized Coefficients (標準化係數)	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) α 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？

答：

- (一)由表2的第一個欄位“未標準化係數”中的B，可知以原始資料表示，從滿意度預測學業成績之直線迴歸方程式為 $\hat{y} = 36.176 + 2.901x$
- (二)以標準分數表示，從滿意度預測學業成績之直線迴歸方程式，則可由表2中之“標準化係數”的Beta，得知為 $Z_{\hat{y}} = 0.726Z_x$ 。表中缺少常數項的標準化係數，是因為將常數項取標準化Z分數之後，其值為0所致。
- (三)設顯著水準為0.05，因針對滿意度的自變項所進行之t考驗，其p-值為0.000(表2)，表示在0.05可容忍犯型I錯誤的機率上限下，該20名學生的樣本資料，提供充分證據足以結論母體迴歸式中的迴歸係數顯著不為0，在得以推翻虛無假設 $H_0: \beta_1 = 0$ 後，足見自變項「滿意度」與依變項「學業成績」之間存在顯著關聯性，因此在子題(一)與(二)中的直線迴歸方程式，在進行解釋與預測時，具有顯著的統計意義。
- (四)若某一國中生對學校滿意度的分數剛好為該20名學生的平均數，由表1可知為12.35，則將其代入迴歸模式，可知理論上該生的學業成績如下：

$$\hat{y} = 36.176 + 2.901x = 36.176 + 2.901(12.35) = 72.00335 \cong 72 \text{ (分)}$$
- (五)若有一國中生對學校滿意度為13分，則將其代入迴歸模式，可知理論上該生的學業成績如下：

$$\hat{y} = 36.176 + 2.901x = 36.176 + 2.901(13) = 73.889 \cong 74 \text{ (分)}$$

【參考書目】

1. 傅立葉老師《測統教材講義》第一回，第八章p.42之十(電腦報表分析)，及練習題與上課筆記。

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-Leaf Plot)。請根據此莖葉圖回答下列問題：(每小題5分，共25分)

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？

答：

- (一)使用莖葉圖來展示資料的優點有：1.快速整理龐雜的資料，確可保留所有原始資料；2.代表不同分組的枝數與其後代表資料筆數的葉數，可以了解資料的分布情形，以取得敘述統計的基本資訊；3.可據以判斷資料的分配型態是否近似對稱的常態分配，或者為偏態；4.可檢視在資料分組過程中，是否遺漏或重複單一筆資料，並加以更正。
(二)依葉的加總可知，此一莖葉圖是由50名學生的分數構成。
(三)此次隨堂考試之最高分與最低分，分別為98分與42分。
(四)此次隨堂考試之眾數與中數分別為70分與72.5分。其中，因為有偶數個觀察值，其中數為第25筆(72分)與第26筆(73分)兩分數之平均數。
(五)依上圖判斷，因中數兩側，以高分組的人數相對較多，在對稱分配下，此次隨堂考試之平均數應大於眾數。

【參考書目】傅立葉老師《測統教材講義》第一回，第二章p.7(莖葉圖優點)，及練習題與上課筆記。

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。(5分)
(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。(10分)
(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？(5分)

(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。(5分)

答：

- (一)效標關聯效度的意義是指，測驗編製者在檢視自己全新編製的測驗、量表，或測量工具之品質時，將參加測驗者的分數，與受試者在某一客觀公正的效標上之表現，進行相關係數的計算與關聯性的探討，例如：新編的學生領導能力測驗，與學生在校園的擔任班級與社團幹部的表現，如果相關性夠高，則提供該新編測驗有高效度之一種證據，即稱效標關聯效度。
- (二)兩種最常用於表現效標關聯效度的類型為：
- 1.同時效度 (concurrent validity)：

測驗分數與選用效標為來自同一時間或相近時間，如前所述，以受測學生該學期在校是否擔任幹部，了解其在領導力測驗的分數高低，所計算而得的效標關聯效度即為同時效度。
 - 2.預測效度 (predict validity)：

當測驗分數與取得效標資料的時間落差相當大，例如：性向測驗所預測個別受試者在未來的發展傾向與領域的結果，與受試者多年後的成就之相關性的探討，即為預測效度。因研究該測驗的時效拉的相當久遠，通常需要龐大的預算與人力支持，進行上較費力。
- (三)因為性向測驗以能夠準確預測受試者未來的性向表現為其最重要的目的，由表中的效度係數可知，除測驗A與C相隔六個月的預測效度係數(0.45與0.55)，並未明顯低於其相隔一個月的預測效度係數(0.5與0.6)；至於測驗B的兩種預測效度係數差距達0.5，且其相隔六個月的預測效度也太低，因此應以「相隔六個月的預測效度」較不合理。
- (四)在刪除前述的「相隔六個月的預測效度」後，三個測驗的優劣排序為BCA，理由仍以性向測驗主要在預測未來的行為表現，在其他條件相同下，三個測驗在內容效度與同時效度的差距不大，此一排序仍以「相隔一個月的預測效度」係數高低為主要考量。

【參考書目】傅立葉老師《測統教材講義》第三回，第十六章p.18-19與上課筆記。