

《教育測驗與統計概要》

- 一、Robert Linn曾在其著作提到效度乃是一種一元論的概念，只是具有多種型式的證據，包括內容、建構、效標關聯和後果等四個面向。Linn也提到就許多測驗與評量而言，既不實際也不需同時考量上述四個面向。試就此四面向給予簡單定義，並以教室班級評量及學術性向測驗為例，說明各所強調的主要考量效度面向為何？較次要的面向為何？（30分）

試題評析	本題涉及特定學者之相關論點，特別提到後果效度之概念，算是有點偏頗。不過在其他三種效度證據之定義上，以及以兩種測驗為例，要求考生判斷何種效度證據為重的答題上都不是太難。考生應能取得不錯的分數。
考點命中	高點《教育測驗與統計講義》第三回第16章，傅立葉編撰，頁23-24。

答：

- (一)一個測驗或評定工具之效度的多種證據，可以包含內容、建構、效標關連與後果等四個面向，其定義分別如下：
- 1.內容效度：指測驗內容(組成題目)的教材與考試範圍之代表性與試題取樣的適切性，又稱為邏輯效度。通常內容效度的證據屬於質化，可以由專家意見(共同認定)或雙向細目表(確保試題涵蓋教學目標與教材內容兩個面向)取得測驗具內容效度之證據。
 - 2.建構效度：指一測驗能夠測量到理論上的構念或特質的程度，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。常須就歷史文獻對於構念的定義進行脈絡回顧與探討，才能掌握心理或人格構念的真實含意，進而編組良好試題組合。
 - 3.效標關聯效度：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，又稱經驗效度(empirical validity)；或因其證據屬於量化指標，又稱統計效度(statistical validity)。
 - 4.後果效度：基於測驗評量對社會影響應該負起的責任，因著重於評量的後果影響面，一般將此稱之為後果效度(consequential basis of validity)，意含除了前述三種效度的證據考量外，評量對學生個人、老師教學、及班級文化、甚或社會習慣的影響，都是評量效度的範圍。
- (二)若以教室班級評量與學術性向測驗為例，因評量目的與評量結果的應用不同，其所強調的主要效度面向不同。教室班級評量應以取得學科的學習成就資訊為主要目的，因此主要考量效度面向以試題出處與組合為探討的內容效度，以及試題與另一外在效標之關聯性高低所提供的間接證據之效標關聯效度為主。而因學習成就的定義不易偏差以及測驗結果頂多影響教師教學的調整，建構效度與後果效度只是次要考慮之效度面向。在學術性向測驗上，因為測驗之性向能力多元，以及測驗結果對於學生升學與就業輔導影響相當大，主要考量效度面向應為建構效度與後果效度。

- 二、利用某一「人格量表」與「教師評定」測量學生三項不同的特質：成就性(甲)、社會性(乙)及獨立性(丙)。根據資料分析所得之多重特質多重方法矩陣如附表。就表中資料，分別指出「相同方法測量相同特質」、「不同方法測量相同特質」、「相同方法測量不同特質」及「不同方法測量不同特質」的係數，並依據係數評論該人格測驗是否具有效度證據？(25分)

附表：人格量表與教師評定之多重特質多重方法矩陣

	人 格 量 表			教 師 評 定		
	甲	乙	丙	甲	乙	丙
人格量表						
成就性 (甲)	.71					
社會性 (乙)	.62	.89				
獨立性 (丙)	.92	.72	.67			

教師評定

成就性 (甲)	.79	.66	.64	.83		
社會性 (乙)	.60	.61	.63	.80	.87	
獨立性 (丙)	.51	.81	.85	.74	.69	.84

試題評析	本題屬於測驗學的構念效度範圍。特別使用多重特質多重方法(MTMM)矩陣將兩測驗工具對於學生三種人格特質測驗結果的相關係數呈現，答題關鍵在於平日練習經驗，以及能否將構念效度成立的兩大效度概念的證據提出。本班同學在英文補充講義已經獲得充分練習，在本題取得高分應無困難。
考點命中	高點《教育測驗與統計補充英文講義》，傅立葉編撰，頁7-8。

答：

(一)依題意，利用兩種量表與評定測量學生的三種特質，結果以多重特質多重方法矩陣呈現如附表。可得四類係數，分別為：

1. 「相同方法測量相同特質」的相關係數，包括：
人格量表的.71, .89, .67，以及
教師評定的.83, .87, .84
2. 「不同方法測量相同特質」的相關係數有.79, .61, .85
3. 「相同方法測量不同特質」的相關係數，包括：
人格量表的.62, .92, .72，以及
教師評定的.80, .74, .69
4. 「不同方法測量不同特質」的相關係數有.60, .51, .81以及.66, .64, .63

(二)依據前述各相關係數的證據，由於「不同方法測量相同特質」的相關係數(.79, .61, .85)夠高，提供收斂效度的證據；以及「不同方法測量不同特質」的相關係數[.60, .51, .81]以及[.66, .64, .63]較低，也提供區別效度的證據，因此該人格測驗(量表)具有構念效度的證據，足以作為學生三種人格特質的測驗、檢視與解釋。

三、研究者想知道學生的數學自我效能對其數學學習表現的預測力，他蒐集30名學生的數學自我效能量表分數(X)以及數學成績(Y)，下表是這兩個變項的描述統計，兩者的相關係數為0.55。

	X	Y
平均值	35	82
標準差	2	5

(一)請寫出以數學自我效能量表分數預測數學成績的非標準化的(unstandardized)迴歸方程式，解釋迴歸係數的意義。(15分)

(二)該迴歸方程式的估計標準誤為何？(5分)

試題評析	本題為最基礎的迴歸分析計算題，除了估計迴歸係數以構建方程式外，考生仍需仔細解釋迴歸係數的意義，並計算估計標準誤。此題難度不高，本班同學能夠100%拿得所有分數，一般考生分數高低的關鍵在於迴歸係數的意義解釋，如果無法強調「平均而言」則無法凸顯迴歸分析所強調的主要精神。
考點命中	高點《教育測驗與統計講義》第一回第8章，傅立葉編撰，頁49-52。

答：

(一)利用題目提供的數學自我效能量表分數(X)與數學成績(Y)兩變項的平均數與標準差，以及相關係數之資訊，可以進行迴歸方程式的兩個參數估計，計算如下：

$$b_1 = r \frac{s_Y}{s_X} = 0.55 \left(\frac{5}{2} \right) = 1.375$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 82 - 1.375(35) = 33.875$$

因此，以數學自我效能量表分數(X)預測數學成績(Y)的非標準化迴歸方程式為

$$\hat{y} = 33.875 + 1.375x$$

方程式中之迴歸係數1.375的意義為，在利用數學自我效能量表分數(X)的資訊以及此一迴歸方程式的情況下，研究者可以清楚知道，每增減數學自我效能量表分數(X)1分，平均而言，數學成績(Y)會隨之增減1.375分。

(二)該迴歸方程式的估計標準誤可以計算如下：

$$s_{Y.X} = 5\sqrt{1 - 0.55^2} = 4.175 \text{ (分)}$$

四、某研究者想知道排行老大的智商是否高於老二，表一是15個家庭老大與老二的智力測驗分數，請以適當的統計方法檢驗研究者的想法。請列出統計假設並詳述其檢驗步驟 ($\alpha = .05$)。(25分) (附t表如表二)

表一

家庭代碼	老大	老二	差異
1	124	114	10
2	115	102	13
3	110	127	-17
4	139	104	35
5	116	91	25
6	88	102	-14
7	120	104	16
8	100	102	-2
9	91	119	-28
10	94	88	6
11	102	119	-17
12	123	132	-9
13	126	114	12
14	105	109	-4
15	102	109	-7
平均值	110.33	109.06	1.27
標準差	14.58	12.14	17.40

表二 t分配的自由度與百分點

df	55	60	65	70	75	80	85	90	95	97.5	99	99.5	99.95
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.127	.256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	.126	.254	.387	.527	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Table D is adapted from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research*, published by Oliver & Boyd Ltd., Edinburgh, and by permission of the authors and publishers.

* The lower percentiles are related to the upper percentiles which are tabulated above by the equation $p_{1-\alpha}^* = -t_{1-\alpha}^*$. Thus, the 10th percentile in the t -distribution with 15df equals the negative of the 90th percentile in the same distribution, i.e., $t_{0.10}^* = -1.341$.

試題評析

本題考的是兩母體平均數差異的推論統計。答題重點在於選擇t-檢定前能否先進行「變異數同質性」條件成立與否的F-檢定，以滿足題目所要求的詳細步驟。題目難度不高，班內考生皆能從容應付答題。

考點命中

高點《教育測驗與統計講義》第二回第9章，傅立葉編撰，頁20。

答：

依題意，本題的假設檢定為單尾檢定，令樣本家庭老大與老二的平均智商分別為 \bar{X}_1 與 \bar{X}_2 ，則統計假設為

$$H_0: \mu_1 \leq \mu_2$$

$$H_1: \mu_1 > \mu_2$$

因為15個家庭的小樣本，且老大與老二兩母體智力測驗分數標準差未知，應先進行「變異數同質性」條件成立與否的F-檢定如下：

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$F = \frac{14.58}{12.14} = 1.201$ ，小於2.95的臨界值，不拒絕虛無假設，「變異數同質性」並未被違反，隨之可採t-檢定。

首先，計算共用的變異數估計值如下：

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{14(14.58^2) + 14(12.14^2)}{15 + 15 - 2} = 179.978$$

$$t = \frac{1.27}{\sqrt{179.978\left(\frac{1}{15} + \frac{1}{15}\right)}} = 0.259$$

，小於表二之1.771的臨界值，不拒絕虛無假設，該研究者無法利用這15個

家庭的樣本資料提供充分證據足以結論老大的智商高於老二。該檢定未達顯著。

高
點
·
高
上

【版權所有，重製必究！】