

《教育測驗與統計》

試題評析

今年試題與往年最大不同處在於，今年著重考生平時對統計與測驗學的理解與實用(活用)，而不再只是死背的能力。第一題考的是「信度指標alpha係數」的應用，上課老師已對本班學員花許多時間講解；第二題為「卡方統計法」的計算與解釋；第三題更是班內同學再熟悉不過的「單因子變異數分析」；第四題則為「試題反應理論」。綜言之，四題分別在班內平日教材的Ch.15, 12, 13, 19中著墨甚多，本班考生平均水準應可得75分的高分。

- 一、有一位部會首長認為良好的情緒管理有助於凝聚團隊向心力，提升工作品質及行政效率，於是他請人事單位諮詢情緒管理專家，草擬乙份「情緒量表」。量表初稿共25題 (item1~item25)，採李克式 (Likert type) 量表法計分 (如最符合4分，最不符合1分)。為了檢定此量表是否具有可接受的信度 (reliability)，於是抽取該部會同仁 (男30，女30) 進行預試 (tryout)，得到初步的結果如表1 (總Cronbach's α 值 = .836)。

表 1

題號	項目刪除時的 Cronbach's α 值
item1	.831
item2	.827
item3	.839
item4	.834
item5	.839
item6	.831
item7	.826
item8	.837
item9	.831
item10	.824
item11	.832
item12	.834
item13	.828
item14	.840
item15	.829
item16	.827
item17	.821
item18	.824
item19	.833
item20	.828
item21	.836
item22	.822
item23	.833
item24	.824
item25	.828

表 2

題號 (略)	項目刪除時的 Cronbach's α 值
item	.853
Item	.848
Item	.853
Item	.846
Item	.845
Item	.851
Item	.846
Item	.853
Item	.855
Item	.847
Item	.849
Item	.849
Item	.839
Item	.845
Item	.852
Item	.848
Item	.840
Item	.853
Item	.845
Item	.846

此資料經專家判斷後，認為其中有五題的數據不盡理想，因此，決定以20題為正式量表。請根據以上敘述及表1的資料，回答下列問題。

- (一) 假設妳/你是專家，則要刪除那五題 (請寫出題號) ? (5分)。

- (二)刪除五題後，再以保留的20題重新估算 α 值，結果專家對項目刪除時的Cronbach's α 值及20題之總Cronbach's α 值皆十分滿意，則此20題量表的信度係數至少會是多少(參考表2)?(5分)
- (三)為了在檢定此情緒量表的實徵性效度(validity)，經查，該部會同仁前兩個月曾做過智力測驗、性向測驗、特殊性向測驗及人格測驗，請問要選擇此四種測驗中那一種測驗的資料與情緒量表分數求相關較為適合?(5分)
- (四)如果正式情緒量表分數與所選擇的測驗之分數其皮爾遜相關係數為.78，請問此為那一種類型的效度係數?(5分)
- (五)正式量表之信度及效度經檢定皆合乎要求後，即對該部會的全體同仁施測。因此，每一位同仁皆會有20題的個別題得分(1~4分)，也會有此20題的總分(20分~80分)。請問妳/你要瞭解同仁的情緒穩定情形時，要以20題逐題得分解讀，或從20題的總分高低來瞭解同仁情緒穩定程度。(分數愈高表示情緒愈穩定；反之，愈不穩定)?(5分)

答：

- (一)利用表一的資訊可知該專家應刪除的題號為3, 5, 8, 14, 21
- (二)0.84
- (三)人格測驗，以人格測驗的五大人格特質來說，其中之一就是所謂的情緒穩定性，與此一情緒量表所欲測量的人格特質有關，可以成為該量表的最佳外在效標。
- (四)0.78是該量表的效標關聯效度。
- (五)要了解同仁情緒穩定性的高低，可以依測驗結果總分的高低來加以了解，不過，視需要仍得利用個別試題的回答情形與得分來取得更進一步的詳細解釋依據。

二、有一位署立醫院醫師正在進行一項婦女懷孕期間抽菸與喝酒對新生兒影響的研究，於是調查其研究樣本(100人)的抽菸與喝酒的習慣，結果如下表之觀察次數。

表：100名婦女懷孕期間抽菸與喝酒之觀察次數與期望次數

			飲酒			總和
			不飲酒	量少	量多	
抽煙	量多	觀察次數(O)	3	9	18	30
		期望次數(E)	12.6	9.3	8.1	30.3
	量少	觀察次數(O)	15	14	6	35
		期望次數(E)	14.7	10.9	9.5	35.0
不抽煙		觀察次數(O)	24	8	3	35
		期望次數(E)	14.7	10.9	9.5	35.0
總和		觀察次數(O)	42	31	27	100
		期望次數(E)	42.0	31.0	27.0	100.0

- (一)請妳/你依據上表的資料，計算卡方(Chi-Square, χ^2)值。(10分)
- (二)要檢定上表兩個變數(抽煙與飲酒)是否獨立性(test of independence)，其卡方的自由度(degrees of freedom; df)是多少?(5分)
- (三)若設顯著水準 α 為0.05，經查卡方分布表(Chi-Square Distribution Table)可得 $\chi^2_{.95, df}$ 為9.488(df如小題二)，那麼能否推斷婦女懷孕期間抽菸與喝酒是否獨立?(10分)

答：

- (一)32.7
- (二)自由度為(3-1)(3-1)=4
- (三)由於小題一的卡方統計量32.7大於9.488的臨界值，起碼就此一百位婦女樣本所代表的母體而言，可以推斷婦女懷孕期間之抽煙與喝酒的行為是不獨立，也就是有關的。

三、有一位主管想瞭解其同仁上班帶飯盒（便當）對家庭滿意度的影響，於是她調查其單位內所有同仁帶飯盒的情形，結果如表1。家庭滿意度問卷得分愈高表示對其家庭愈滿意。為了探討中午「帶飯盒」、「偶爾帶飯盒」及「不帶飯盒」的同仁對其家庭滿意度的差異情形，因此，將調查所得之資料進行變異數分析，結果如表2。表2資料不慎在處理時刪除了大部分的數據，現在請妳／你想辦法將表2中的數據恢復，並請依據以上敘述回答下列問題。

表1：家庭滿意度之描述統計量

組別	人數	平均數	標準差	標準誤
不帶飯盒	32	31.9688	1.51305	.26747
偶爾帶飯盒	26	34.1923	.93890	.18413
帶飯盒	20	36.4500	.99868	.22331
總和	78	33.8590	2.17258	.24600

表2：家庭滿意度之變異數分析摘要

變異來源	平方和	自由度	均方	F	顯著性
組間	?	?	?		
組內	111.957	?	?	?	.000
總和	363.449	?			

(一)為了檢定此三組對家庭滿意度的差異情形，請問妳／你如何陳述其研究假設？(5分)

(二)請妳／你找出表2中之F值。(10分)

(三)若設顯著水準 $\alpha = .05$ ，依據表2中顯著性數據，妳／你如何對分析的結果下結論？(5分)

(四)如果此三組對家庭滿意度有顯著差異，下一步如何處理？妳／你毋需計算，請以文字說明即可。(5分)

答：

首先將表格恢復如下：

變異來源	平方和	自由度	均方	F	顯著性
組間	251.542	2	125.771		
組內	111.957	75	1.493	84.24	.000
總和	363.499	77			

(一)研究假設為：三組不同帶飯盒的家庭，其對家庭滿意度有顯著的影響。

(二)F值為兩個均方的比值，其為84.24。

(三)因為0.05的顯著水準大於0.000的p-值，此一研究問題的變異數分析結果可以拒絕虛無假設，結論：三組不同帶飯盒的家庭，其對家庭滿意度確實有顯著的影響。

(四)因為三組不同帶飯盒的家庭，其對家庭滿意度的平均數有顯著差異存在，應進一步進行事後比較，以找出確實的差異發生在哪些分組之間。

四、項目反應理論又稱試題反應理論 (item response theory; IRT)，早期稱為潛在特質理論 (latent trait theory)，此理論已廣泛應用於教育與心理測驗或其他領域。IRT是現代心理計量學重要的一支，此理論有兩個重要的基本假定 (assumptions)，即單維性 (unidimensionality) 與局部獨立 (local independence)。

(一)請說明單維性 (unidimensionality) 與局部獨立 (local independence) 的意義。(10分)

(二)項目反應理論有多種模式 (models)，目前常被應用者有單參數 (one-parameter logistic或Rasch model) 模式、雙參數 (two-parameter logistic或Birnbaum model) 模式及三參數 (three-parameter logistic或guessing model) 模式。請說明三參數模式中三個參數 (a、b及c) 的意義。(15分)

答：

- (一)單維性指的是測驗本身所欲測量的特質或能力，理論上應只有一種，以免混淆或測量到不相干，甚至是錯誤的特質或能力；至於局部獨立性則指稱測驗中得試題與試題之間，其雖傾向於在測量同一特質或能力，但是在某一題的答對與答錯與否，並不影響其他題目的作答對錯。
- (二)三個參數指的是難度、鑑別度與猜測度。難度是試題本身的困難程度，通常也指考生答對的百分比；鑑別度則為試題鑑別考生能力或特質高低之能力；猜測度則為考生猜測的可能性。三個參數可以決定考生在該題正確回答的機率，且其組合可以決定不同的試題反應模式。