

# 《教育測驗與統計》

- 一、某能力測驗適用對象為小學四到六年級學生，測驗有兩個平行版本，甲式及乙式。測驗發展者抽取100名四到六年級的學生作為信度研究的對象。這100名學生同時完成甲、乙式測驗，然後求取兩式測驗成績的相關，其值為0.8。今有汪老師以此測驗測量小學四年級學生，為確保測驗結果的信度，抽取100名四年級學生，同時完成甲、乙兩式的測驗。
- (一)測驗發展者估計的是何種信度？並解釋信度0.8的意義。(10分)
- (二)請問汪老師所得的信度是否會高於原測驗信度？為什麼？(10分)
- (三)測驗發展者覺得信度不夠理想，打算將信度提升到0.9，如果以增加平方試題方式，測驗長度要如何調整？(5分)

試題評析	本題針對測驗的複本信度命題。並針對受試者同質性與提高信度等在子題中考驗考生對於影響該信度之因素所造成的結果。題目難度不高，考生應能容易取分。
考點命中	《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，第十五章，頁12。

答：

- (一)依題目敘述，該能力測驗的發展者利用100位四到六年級學生的研究對象，所計算甲式與乙式兩個平行版本測驗成績的相關係數0.8，其估計的是複本信度係數。其值表示100位學生樣本在兩個複本之測驗成績的相關程度很高，意即，學生在甲式測驗的得分愈高，則在乙式測驗的得分也會較高；反之，若在甲式測驗的表現較差，則在乙式測驗的表現也會較差。兩個題本得分的高度相關提供了該測驗高信度的證據。
- (二)汪老師針對100位小學四年級學生施測兩個題本，利用測驗結果所計算之相關係數所代表的複本信度應該會高於原測驗的信度，因為在相等樣本量的情況下，相對於原測驗發展階段所取得之信度係數所使用的四到六年級小學生，汪老師所施測的學生因同為小學四年級學生，其同質性較高而會有較高的信度。
- (三)若要將測驗信度由0.8提高到0.9，通常以增加題目數量為主要方法，但是增加的題目應與原有題目就題型、測量特質能力、難度等面向具備同質性的條件，而且增加題目應以緩步增加的方式，逐步試算測驗的信度，同時進行必要的題目刪減動作。並必須避免過度增加題數，以免反而降低測驗信度。

- 二、隨著教育評量的改革趨勢，建構式反應(constructed-response)題型/實作評量已普遍出現於各種考試，但部分學者、教師、家長、學生質疑此種題型的評分過於主觀，可能造成偏誤，因而對此種形式的試題持保留態度或反對立場。請就你對這類評量的瞭解，及它在班級評量的應用情形，至少舉出三種措施使此種題型或評量的偏誤最小化。(15分)

試題評析	本題涉及實作評量所面對最為人詬病，也是實務工作者最感困難的評量偏誤進行命題。題目要求考生提出降低評量偏誤的三種可能措施並無一定標準答案，只要從評量目的、內容、方式以及評分計分形式加以回答即可獲得閱卷者的青睞。雖然命題不意外，但是大多數考生準備階段不太會鎖定此一本班講義一再強調之重點。
考點命中	1.《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，第十九章，頁49-52。 2.《高點·高上教育測驗與統計考前重點整理》，傅立葉編撰，頁1-2、10-11。

答：

【版權所有，重製必究！】

- (一)實作評量稱不上是新的評量方式，近來之所以受到注意的主要原因是美國教育界廣泛將其運用在學校學科領域中，並且做為評估學校教學績效和瞭解學生成就水準一般狀況的策略。此一評量方法受到歡迎的主要原因之一是：一般人對於選擇式的測驗題感到不滿意，例如，選擇式的測驗題只能測量學生「知道」什麼，但卻無法測量學生「能做」什麼；此外，認為以選擇題為主的標準化測驗，在教師教學績效的壓力之下，造成部分教師教學側重於測驗內容，而扭曲了教學的面貌和窄化學生的學習結果。此外，目前一些入學考試和證照考試也在原有的選擇題之外，增加建構反應題或實作測驗部分，實作評量在教育界和測驗界已是一個非常顯明的趨勢。

(二)以觀察和專業判斷來評量學生學習成就的評量方式都可以稱為實作評量，其型式非常多元化，例如建構反應題、書面報告、作文、演說、操作、實驗、資料蒐集、作品展示等，都是實作評量的例子，案卷評量也是實作評量的一種型式。實作評量除了需要時間和人力去計分外，評量和觀察重點的掌握和評分標準的訂定有時也是一個難題，尤其是對非結構性的作業項目進行評量。如何使實作評量的偏誤最小化，建議以下三種措施：

1.判斷的規準(criteria)和標準(standard)：

應事先明確列出行為表現的重要層面和各層面表現的評分標準，例如怎樣的表現是優異的、普通的或是不佳的。該項實作評量品質關鍵的所在，教師要事先和學生就這一項進行溝通，給學生回饋時也是從這裡切入。

2.計分型式的選擇：整體性計分或分析性計分？

當只是一般性的決定，如分組、評選，由於只須單一的整體分數，適合採用整體評分法；如果是具有診斷困難、或瞭解學生精熟表現水準之功能的決定，則使用分析評分法。

3.評量目的、評量內容與評量方式的最佳配適：

教師可以根據教學目標和評量目的，編擬適用的評量計劃。並考慮要評量的行為表現是什麼？觀察行為表現的重點是著重於「歷程」(process)或是「作品」(product)，或是二者兼顧？然後挑選適合的評量工具。

三、某房屋仲介公司的網站，定期公告每個經紀人售出房價的集中量數，一般來說，每個房屋經紀人所賣出的房屋中，會有少數特別昂貴，不過每個人所賣出的房屋數相差無幾。網站上也呈現該地區兩個地段A與B的房價標準差，這個月A地段售出的房子有20戶，標準差為12萬元/坪，B地段售出的房子有30戶，標準差為8萬元/坪。

(一)若房屋經紀人的佣金為賣出總房價的5%，請問那個集中量數與房屋經紀人的收入最相關？為什麼？(5分)

(二)王先生想利用該網站粗估一下該地區的房屋售價散布情形，請問將兩個地段資料合併後，該地區的房價標準差是多少？(6分)

(三)若該仲介公司改變佣金制度，將佣金從5%減少為4%，請逐一說明房屋經紀人的收入之變異量數分別有何影響？(1)全距 (2)四分差 (3)變異數。(9分)

試題評析	本題以敘述統計之統計量數為命題重心。其中，又以變異量數的分數為重。因佣金比例降低所導致三種變異量數之變化，以及合併變異數的計算都只是運算性質與簡單計算公式的應用，一般考生應能掌握本題得分。
考點命中	1.《高點·高上教育測驗與統計講義》第一回，傅立葉編撰，第三、四章，頁12-13、17-18。 2.《高點·高上教育測驗與統計考前重點整理》，傅立葉編撰，頁1、10。

**答：**

(一)集中量數常用的有平均數、中位數與眾數等三個。因仲介房屋公司網站定期公告每一經紀人售出房價的集中量數，而經紀人的佣金為售出總房價的5%，因為總房價的計算只能用平均售出房價乘上售出房屋數取得，因此，平均數與房屋經紀人的收入最為相關。

(二)合併A與B兩地段房屋售價的資料後，該區房價的標準差可以計算如下：

$$s = \sqrt{\frac{(20-1)(12^2) + (30-1)(8^2)}{20+30-2}} = \sqrt{\frac{4592}{48}} = 9.78(\text{元})$$

(三)當佣金制度改變，由5%降至4%，其對於房屋經紀人收入的變異量數所發生影響依序說明如下：

1.全距：其計算式為最高與最低佣金收入相減的差值，因佣金比例減少，全距將因最高與最低佣金縮水而變小。

2.四分差：其計算式為第三四分位數與第一四分位數相減後除以二，因佣金比例減少，四分差將因兩個四分位數的佣金金額縮水而變小。

3.變異數：其計算利用所有房屋物件售價與平均售價的離均差平方和除以房屋件數，因為佣金比例減少，變異數將因所有售出房屋每筆佣金金額縮水而變小。

四、王老師目前任教小學三年級，他想瞭解班上同學的體重會不會有過重的情形。在全班所有30名小朋友量過體重後，王老師想找個數據來描述小朋友體重資料的集中情形而來請教你。請問你如何向王老師說明平均數、眾數、中位數三種集中量數指標的意義？（15分）

試題評析	本題針對三種集中量數，以探討任教班級學生體重集中情形的王老師為對象，要求考生對其提供每一種統計量數之意義說明。難度不高。
考點命中	1.《高點·高上教育測驗與統計講義》第一回，傅立葉編撰，第三章，頁12-13。 2.《高點·高上教育測驗與統計考前重點整理》，傅立葉編撰，頁1、3。

**答：**

依題意敘述，王老師想要探討任教之小學三年級班上學的體重集中情形。身為王老師請教的對象，三種集中量數指標分別說明其意義如下：

(一)平均數：

計算時利用所有學生的體重資料，雖然簡單易懂，但容易受極端值的影響而產生低估或高估；因此，若確定班上同學存在過重或過輕的個人時，應建議王老師改採中位數進行體重集中情形的說明。

(二)眾數：

以出現次數或頻率最多或最高的學生體重作為解釋全班同學體重的集中情形，困擾之一是因眾數不一定存在而無法讓王老師順利進行解釋；另外，當眾數不只一個時，也會讓王老師在解釋時產生困擾。

(三)中位數：

當王老師將30位班上同學的體重依大小排列，則體重數據位於正中間者可用做全班同學體重集中情形的解釋依據。這是前述當存在體重過重或過輕的離群值時，用以取代平均數的集中量數。但是小小的缺點是僅使用一位學生的體重數據做為全班體重集中情形的量測，並無法充分利用所有體重測量值。

五、研究調查不同學院的大學生之打工經驗滿意度是否不同，在文、理、工學院各抽取100名具打工經驗的學生，其中對打工經驗表示滿意的分別各有70、75、78名，其餘表示不滿意。

(一)請利用統計檢定不同學院的學生其打工經驗滿意度是否不同？請說明檢定須查何表？如何在表中找到臨界值A？並說明如何利用臨界值A做決策。（20分）

(二)這個假設檢定的基本假定 (assumptions) 為何？（5分）

試題評析	本題是今年考試唯一一題推論統計的計算題。考的是類別資料的列聯表之同質性卡方檢定。
考點命中	1.《高點·高上教育測驗與統計講義》第二回，傅立葉編撰，第十二章，頁35-38。 2.《高點·高上教育測驗與統計考前重點整理》，傅立葉編撰，頁1、5。

**答：**

(一)依題意，在調查不同學院大學生對於打工經驗的滿意度研究中，若要檢定不同學院的滿意度是否不同，則因文、理、工三個學院屬於名目(類別)變項，而滿意度以滿意與不滿意進行人數統計為二分名目變項，因此本研究應使用卡方統計法進行百分比同質性的卡方檢定。調查結果可製成表格如下：

	文	理	工	總計
滿意	70 (74.33)	75 (74.33)	78 (74.33)	223
不滿意	30 (25.67)	25 (25.67)	22 (25.67)	77
總計	100	100	100	300

虛無與對立假設為：

$$H_0: p_1 = p_2 = p_3$$

$$H_1: p_j \text{ 不全等}$$

先計算細格觀察人數的期望人數(括號內數據)後，檢定統計量，也就是卡方值計算得：

$$\chi^2 = \frac{(4.33)^2 + (0.67)^2 + (3.67)^2}{74.33} + \frac{(4.33)^2 + (0.67)^2 + (3.67)^2}{25.67} = 1.712$$

$\chi^2$ 查表值為  $\chi^2_{0.95,2} = 5.99$

因為1.712的檢定統計量落在接受域，檢定結論為不拒絕虛無假設，利用這300位不同學院大學生的樣本資料無法提供充分證據足以結論：不同學院學生對於打工經驗的滿意度存在顯著差異。

此一檢定過程需要查卡方表，利用0.05的顯著水準與特屬於列聯表的(3-1)(2-1)自由度，可以找到臨界值A，當檢定統計量因大於臨界值而落在右邊的拒絕域，將獲得拒絕虛無假設的結論；而上述因為檢定統計量因小於臨界值而落在左邊的接受域，所以獲致不拒絕虛無假設的檢定結論。

(二)此一檢定的基本假設為：

1.對於細格期望人數的要求：

一般對於卡方檢驗中的每一個細格，要求其最小期望人數均大於1，且至少有4/5的細格期望人數大於5，此時卡方統計量的計算結果才是可用的。

2.需控制邊際人數，樣本的人數配置應盡量依符合各個群體佔母體總數的比例來抽樣，以求樣本具有高的母體代表性，而使檢定結果的考驗力夠高。

【版權所有，重製必究！】