機率統計考題探討

黃文璋
國立高雄大學應用數學系

98年1月的**學測**(學科能力測驗), 及當年7月的**指考**(大學入學指定科目考試), 在數學科中, 開始出現**信賴區間**的考題。這是由於**九五數學課綱**引入信賴區間等統計題材, 而95年入學的高中生, 乃98年畢業。

修訂的**九九數學課綱**, 雖刪除**交叉分析**此一不宜出現在高中數學的題材, 卻仍保留信賴區間的部分。眾所周知, 考試引導教學乃不可避免。因此我們好奇, 自這些不算容易的統計概念進入高中數學後, 到底學測及指考, 機率統計考些什麼? 從大考中心網站, 找到從98至101年四年的學測, 及指考數學甲、數學乙, 共12份考題, 大致翻閱一遍後, 這才知道現今高中數學, 機率統計被重視的程度, 超乎我們想像。以98年指考數學乙為例, 總共12道題目裡, 機率統計佔了7題, 共49分, 幾乎佔總分100之半。由於高中數學課本, 對信賴區間這塊, 往往寫得零零落落, 那考題又如何? 不禁令人滋生疑竇。本文便是從前述12份考題中, 挑出幾個題目來跟大家討論。題目皆一字未改, 不過在討論時, 對於"分配"一詞, 我們以"分佈"取代。

1. 98年學測數學第9題(多選)

某廠商委託民調機構在甲、乙兩地調查聽過某項產品的居民佔當地居民之百分比(以下簡稱為「知名度」)。結果如下: 在95%信心水準之下, 該產品在甲、乙兩地的知名度之信賴區間分別為$[0.50, 0.58]$、$[0.08, 0.16]$。試問下列哪些選項是正確的?

(1) 甲地本次的參訪者中, 54%的人聽過該產品

(2) 此次民調在乙地的參訪人數少於在甲地的參訪人數

(3) 此次調查結果可解讀為: 甲地全體居民中有一半以上的人聽過該產品的機率大於95%

(4) 若在乙地以同樣方式進行多次民調, 所得知名度有95%的機會落在區間$[0.08, 0.16]$

(5) 經密集廣告宣傳後, 在乙地再次進行民調, 並增加參訪人數達原人數的四倍, 則在95%信心水準之下該產品的知名度之信賴區間寬度會減半(即0.04)

大考中心公佈的答案為(1), (2)。

註. 大考中心在試卷中附有參考公式及可能用到的數值。其中有一條

95%信心水準下之信賴區間: $[\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$。

討論. 選項(3)正確與否, 頗值得斟酌。首先, 什麼叫"解讀"? 解讀是一件很主觀的事, 不見得須很科學。每當有一颱風接近台灣, 國內外各氣象專家, 對同樣的數據, 可有差異極大的解讀, 即為一例。更不要說對於較政治性的事件, 不同立場的人, 其解讀常是南轅北轍。若以p表甲地全體居民中聽過該產品之比率, 選項(3)即問此次調查結果可否解讀為$p \geq 1/2$的機率大於0.95。依黃文璋(2011)一

文第7節之說明, 雖p並非隨機變數, 但認為$p \in [0.50, 0.58]$之機率為0.95, 可視為對機率主觀的解釋, 乃有其道理。而由此即得$p \geq 0.5$的機率大於0.95之主觀機率。但考對機率的主觀解釋卻是不宜的。例如, 考慮如下題目: 下列那一隊得下屆世界盃足球賽冠軍之機率最高?

(1)法國隊, (2)巴西隊, (3)德國隊, (4)英國隊, (5)荷蘭隊。

不論過去資料如何, 這種題目就是不宜考。因涉及主觀機率。即使待全部比賽結束, 冠軍產生後, 你仍無法說當初選那一隊是錯的。因就算眞有那隊得冠軍之機率最高, 也有可能分組便被淘汰, 所謂球是圓的。統計裡的"推論", 與數學裡的"推論"並不相同。統計裡允許各種推論法, 即使投擲一銅板, 連續得10次正面, 都可仍堅持該銅板為公正。由於除了可有不同的主觀機率之原因外, 再加上允許對抽樣結果可有不同的"解讀"及"推論", 故並不宜有選項(3)。而一旦有選項(3), 則便不能說此選項不正確。

2. 99年學測數學第12題(多選)

想要了解台灣的公民對某議題支持的程度所作的抽樣調查, 依性別區分, 所得結果如下表:

	女性公民	男性公民
贊成此議題的比例$\hat{p}$	0.52	0.59
$\hat{p}$的標準差$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	0.02	0.04

請問從此次抽樣結果可以得到下列哪些推論?

(1) 全台灣男性公民贊成此議題的比例大於女性公民贊成此議題的比例

(2) 在95%的信心水準之下, 全台灣女性公民贊成此議題之

比例的信賴區間為$[0.48, 0.56]$(計算到小數點後第二位, 以下四捨五入)

(3) 此次抽樣的女性公民數少於男性公民數

(4) 如果不區分性別, 此次抽樣贊成此議題的比例$\hat{p}$介於0.52與0.59之間

(5) 如果不區分性別, 此次抽樣$\hat{p}$的標準差$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$介於0.02與0.04之間

大考中心公佈的答案為(2), (4) 。

討論. 選項(2)最後括號中的文字為多餘, 因依所給公式及數據即得$[0.52 - 2 \cdot 0.02, 0.52 + 2 \cdot 0.02] = [0.48, 0.56]$, 根本沒有四捨五入的問題。但就當做是企圖迷惑考生, 只好接受。選項(1)則類如上題之說明乃有些問題。選項(2)-(5)皆可明確算出數值。調查結果, 雖男性贊成比例0.59大於女性贊成比例0.52, 但兩者95%信賴區間分別為$[0.51, 0.67]$及$[0.48, 0.56]$, 兩區間有重疊, 這可能是選項(1)不被大考中心視為正確之原因。但在統計裡, 有何理論基礎要求須依信賴區間來做推論? 採用點估計, 因$0.59 > 0.52$, 因而得到(1)之推論不行嗎? 甚至, 即使考慮信賴區間, 又豈有規定必得95%? 68%不行嗎? 此時二信賴區間各為$[0.55, 0.63]$及$[0.50, 0.54]$便不重疊了。再度強調, 在統計裡, 可有各種推論法, 且往往沒有那一推論法是永遠最佳, 只能依不同的標準評比。

3. 98年指考數學甲選擇題第7題(多選)

已知丟某枚銅板, 其出現正面的機率為p, 出現反面的機率為$1-p$, 將此枚銅板丟擲n次, 在丟擲過程中, 正面第一次出現時, 可得獎金1元, 正面第二次出現時, 可再得獎金2元, 正面第三次出現時, 可再得獎金3元, 以此類推。試問下列哪些選項是正確的?

(1) 若n次丟擲中出現正面k次, 總共得到獎金$\frac{1}{2}(k^2-k)$元

(2) 丟擲銅板第二次之後, 累計得獎金1元的機率為$2(p-p^2)$

(3) 總共得到獎金2元的機率為$\frac{n(n-1)}{2}p^2(1-p)^{n-2}$

(4) 總共得到獎金$\frac{1}{2}(n^2-n)$元的機率為$n(p^{n-1}-p^n)$

大考中心公佈的答案為(2), (4) 。

討論. 題目中首句宜說將此銅板"獨立地"丟擲n次, 才較清楚 。有些事件是不獨立的 。如高中數學課本中亦出現的簡單隨機抽樣(取樣後不放回), 各次取樣的結果便不獨立 。不過不少人常省略強調各次試驗為獨立 。只是若各次銅板的投擲不為獨立, 便根本無從判斷各選項正確與否 。在這種大型考試, 題意應更明確, 不宜假設學生都懂命題者的心 。另外, 選項(2)題意有些含混 。何謂丟擲銅板第二次之後, 累計得獎金1元? 是後面$n-2$次累計得1元?

4. 98年指考數學乙選擇題第4題(多選)

國一學生30萬人, 智商測驗的結果是「平均數100, 標準差15」的常態分配 。若以智商130以上做為甄選國一學生為資優生的門檻, 則根據這次測驗的結果判斷下列選項中的敘述, 哪些是正確的?

(1) 約有5%的國一學生通過資優生甄選門檻

(2) 約有15萬名國一學生的智商在100以上

(3) 超過20萬名國一學生智商介於85至115之間

(4) 隨機抽出1000名國一學生, 可期望有25名資優生

(5) 如果某偏遠學校只有14名的國一學生, 那麼該校不會有資優生

大考中心公佈的答案為(2), (3), (4) 。

討論. 這題該斟酌的是文字 。如果問投擲一公正骰子, “可期望”得到點數3.5是否正確? 你覺得如何? “期望值”之意義明確, 但是否等同於“可期望”, 就不明確了 。“可期望”一詞, 乃生活用語, 並非機率裡的專有名詞 。對打的兩支球隊, 在比賽前, 都“期望”會贏, 有何不同? 另外, 智商測驗成績為“離散值”(頂多到小數第2位吧), 常態分佈則為連續型, 離散型怎會變成連續型呢? 問題還不只於此, 即使人數多達30萬, 測驗的結果, 最多只能說可以常態分佈當近似的模型, 而絕非就是常態分佈 。依30萬人之測驗結果, 來建立智商模型可以(也是一種近似), 但說測驗結果有何分佈, 就不是那麼恰當了 。

我們取一本常用的初等機率統計書Walpole *et al.*(2011)p.187中, 對此情況的寫法, 供大家參考。注意所採用的說法, 皆是 “approximately” normally distributed(近似常態分佈), 這可能較妥當:

(i) If the wages are approximately normally distributed and paid to the nearest cent, $\cdots$.

(ii) The weights of a large number of miniature poodles are approximately normally distributed with a mean of 8 kilograms and a standard deviation of 0.9 kilograms.

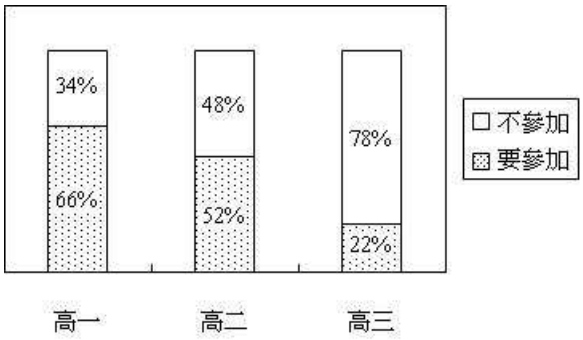
(iii) The IQs of 600 applicants to a certain college are approximately normally distributed with a mean of 115 and a standard deviation of 12.

本題除上述缺失外, 選項(3)其實尚有一無法忽視的問題 。高中數學裡引進常態分佈, 乃為了教信賴區間, 且大多將信心水準取為95% 。如果查標準常態分佈的機率值表, 會得介於正負1.96之間的機率約為0.95 。高中數學課表為了簡便, 將1.96以2替代 。由此導

致對常態分佈, 取值較平均數高兩個標準差以上的機率, 以0.025(=0.05/2)視之 。實際上高兩個標準差以上的機率約為0.0228 。所以隨機抽取1000名國一學生, 其中資優生人數之“期望值”約為22.8, 而非25 。因此題目既然是問何者“正確”, 就算選項(4)“可期望”的說法換成“期望值”, 25也絕非“正確”資優生人數的期望值 。

5. 98年指考數學乙選擇題第6題(多選)

某縣市教育局欲瞭解高中生參加課外活動社團的意願, 開學日隨機調查高一 、高二 、高三學生各1067名, 詢問本學期是否要參加課外活動社團 。已知該縣市的高一 、高二 、高三學生人數幾乎一樣多, 各年級學生調查結果如下圖:



試問下列選項中的敘述, 哪些是正確的?

(1) 學生要參加課外活動社團之比例隨著年級增加而遞減

(2) 由上述資訊可以估算全體學生要參加課外活動社團的比例

(3) 在95%信心水準下, 每一個年級學生要參加課外活動社團的比例之信賴區間, 都可以由題目中已知的數據算出

(4) 在95%信心水準下, 三個年級的調查結果, 以高一學生要參加課外活動社團的比例的信賴區間最長

(5) 在95%信心水準下, 三個年級的調查結果, 以高三學生要參加課外活動社團的比例的信賴區間最短

大考中心公佈的答案為(1), (2), (3), (5) 。

討論. 答案中包含選項(1)有誤。如果大考中心認為第2題的選項(1)不對, 則何以認為本題選項(1)正確? 何況第2題是問可得到那些推論, 這是很緩和的問法, 有如若問你"喜歡"那一個歌手? 則不論說誰, 都不能被認為是錯的。而本題是問那些"正確"? 再強調一遍, 在統計裡, 對同一觀測結果, 每人可有不同的推論。但何者正確, 卻不能僅憑主觀, 或僅憑統計式的推論。選項(1)乃針對母體發問, 而由僅取部分樣本的結果, 如何知母體之眞相究竟為何?

6. 98年指考數學乙選塡題B

某實驗室欲評估血液偵測老年癡呆症技術的誤判率(即偵測錯誤的機率)。共有760人接受此血液偵測技術實驗, 實驗前已知樣本中有735人未患老年癡呆症。實驗後, 血液偵測判斷為未患老年癡呆症者有665人, 其中眞正未患老年癡呆症有660人。試問此血液偵測技術的誤判率為$\frac{(9)}{\underline{(10)(11)}}$。(化成最簡分數)

大考中心公佈的答案為(9): 2, (10): 1, (11): 9 。即誤判率為$\frac{2}{19}$。

討論. 由答案推測此題作法如下:

以"患"及"未患"分別表患老年癡呆症, 及未患老年癡呆症。依題意760人中, 有735人未患, 因此有25人患。偵測結果, 判斷為未患者有665人, 其中眞正未患者有600人。因此未患的735人中, 有

$$735 - 660 = 75(\text{人})$$

被誤判為患; 而有

$$665 - 660 = 5(\text{人})$$

患卻被誤判為未患。誤判共有

$$75+5=80(\text{人})\text{。}$$

因此誤判率為

$$\frac{80}{760}=\frac{2}{19},$$

此即大考中心的答案。只是若依此解法, 則未患及患者之誤判率, 分別為

$$\frac{75}{735}\doteq 0.1020,\ \text{及}\frac{5}{25}=0.2\text{。}$$

患者之誤判率約為未患者之2倍。這可能有異於我們一般所理解的誤判率。

當初接受實驗者, 若全為未患, 則所求之誤判率將為0.1020; 若全為患, 則所求之誤判率將為0.2, 一項偵測之所謂誤判率(還特別括號說是偵測錯誤的"機率")浮動如此大, 不有些奇怪? 不妨這樣想, 體重計標示的誤差, 若與量測者之體重有關, 則該誤差恐怕便沒什麼參考價值。

事實上, $80/760=2/19$, 表那特定的760個接受實驗者, 被誤判的比率。這類的將二數相除, 總可得一比率。但要將該比率詮釋成機率, 就得很謹慎了。此處該比率, 乃自那760人中, 隨機地取1人, 會取中屬於偵測錯誤那群人的機率。此機率, 與該項技術偵測錯誤的機率, 意義不同。又除非接受實驗的760人為隨機產生的樣本, 或至少735:25可視為母體中未患與患的比例, 否則2/19也不能當作自母體中隨機地取1人, 其偵測會有誤之機率的估計值。

底下我們給一例(見黃文璋(2010)p.37例9.4), 這類例子在有關條件機率中常出現, 以說明不論有病或沒病(患或未患), 檢驗可靠度相同(因此誤判率亦相同), 才較合理。

例. 衛生局至中山大學免費檢驗某疾病, 此檢驗之可靠度為90%。即若眞有病, 則有0.9之機率此檢驗呈正的反應(假設只有正負二反應);

若沒有病, 則有0.9的機率檢驗呈負的反應。過去資料顯示平均每5,000人有1人患有此病。此檢驗迅速且無害, 但若檢驗呈正的反應, 則"須"至醫院住院一周做進一步的檢查。問你是否願意接受此檢驗?

解. 以"正"表正的反應,則

$$\begin{aligned} P(\text{有病}|\text{正}) &= \frac{P(\text{正}|\text{有病})P(\text{有病})}{P(\text{正}|\text{有病})P(\text{有病}) + P(\text{正}|\text{無病})P(\text{無病})} \\ &= \frac{0.9(1/5{,}000)}{0.9(1/5{,}000) + 0.1(4{,}999/5{,}000)} \\ &= \frac{9}{5{,}008} \doteq 0.0018 \text{。} \end{aligned}$$

註. 在九九數學課綱裡的"說明與範例", 於第二冊"條件機率與貝氏定理"中, 有底下一段:

> 某一檢查方法對檢驗某一疾病有90%的準確率, 也就是說, 如果患有該疾病的人做檢查, 那麼有90%的機會會呈現陽性反應; 如果沒有該疾病的人做檢查, 也有90%的機會會呈現陰性反應。假設已知全國人口中有2%的人得患有該疾病, 如果有一人以此檢查方法的檢查結果為陽性, 那麼他罹患該病的機率為何?

其中的準確率, 與上例中可靠度之意義相同, 這其實是一般的看法。誤判率與準確率, 也該有類似的定義。因此不理解何以本題對誤判率給出這樣不合理的定義。

7. 99年指考數學乙選擇題第2題(單選)

> 某校高三學生在一次考試中, 成績呈常態分配, 且已知其分數之平均數為70分, 標準差為10分。若從這次考試的學生中, 隨機抽出一位學生, 則這位學生的成績低於60分的機率最接近以下哪一選項?

(1) 0.16

(2) 0.32

(3) 0.34

(4) 0.68

(5) 0.84

大考中心公佈的答案為(1)。

討論. 如第4題, 題目中說"成績是常態分配"不妥。成績取值0至100, 且通常為整數值, 至多到小數第二位。而常態分佈取值任意實數皆可能, 成績並不會有常態分佈, 只能以常態分佈來"近似"。

8. 100年指考數學乙選填題B

為講解信賴區間與信心水準, 數學老師請全班40位同學使用老師提供的亂數表模擬投擲均勻銅板16次。模擬的過程如下: 隨機指定給每位同學亂數表的某一列, 該列從左到右有16個數字; 如果數字為0,1,2,3,4時, 對應投擲銅板得到正面; 而數字為5,6,7,8,9時, 對應投擲得到反面。某同學拿到的一列數字依序為:

$$0612 \quad 9683 \quad 4251 \quad 9138$$

該同學計算銅板出現正面的機率在95%信心水準下的信賴區間: $[\hat{p}-2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p}+2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$。則該同學所得到的結果中, $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \underline{\frac{(9)\sqrt{(10)}}{32}}$。(化為最簡根式)

大考中心公佈的答案為(9):3, (10):7, 即$2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{3\sqrt{7}}{32}$。

討論. 這題倒沒問題。所給16個數字中, 有9個為0,1,2,3,4, 故$\hat{p} = 9/16$。將$\hat{p}$代入所給公式即得。只是這種考題, 豈有任何統計味道?

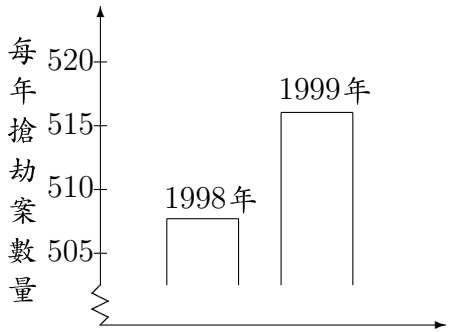
而才第三年(民國100年), 看來信賴區間快沒題目可考了 。

最後我們來看PISA(the Programme for International Student Assessment, 國際學生能力評量計畫, 台灣受測學生包含國中 、五專及高中職)2006數學樣本試題M179, 這是一道統計題目 。

9. 問題: 搶劫

電視主播呈現了下圖並報導:

「從圖表顯示, 從1998年到1999年搶劫案數量有巨幅的上升」。



你認為這位主播對於上圖的解釋是否合理? 請寫出一個理由來支持你的答案 。

滿分

代號21: 不, 不合理 。指出我們看到的只是整個圖表的其中一小部分 。

- 不合理, 須顯示整個圖表 。
- 我不認為那是合理的詮釋, 因為如果顯示全圖的話, 便能看到搶劫案的數目只是輕微上升 。
- 不合理, 因為他只用了圖表上方的小部分 。如果看到全圖由0到520的情況, 便知道上升幅度不是那麼大 。

- 不, 那只是因為該圖表讓人覺得數字巨幅上升。看數字增加並不多。

討論. 到底搶劫案上升能不能算巨幅, 並非只看增加的搶劫案數值之大小, 也宜看發生機率之大小。以男子100公尺短跑的世界紀錄為例。1968年紀錄為9.95秒, 2009年則為9.58秒, 每10年平均快不到0.1秒。因此若在某次比賽中, 某選手將世界紀錄推進0.2秒, 雖只快了約2%, 媒體很可能將大肆報導, 認為是"巨幅"進步。針對本題, 我們給一情境如下。

假設每年搶劫案之數量X有$\mathcal{B}(520, 0.977)$分佈。令$n = 520$, $p = 0.977$。則

$$E(X) = np = 508.04,\ \sqrt{\mathrm{Var}(X)} = \sqrt{np(1-p)} \doteq 3.418\ 。$$

因np及$n(1-p)$皆大於5, 由常態近似, 得

$$\begin{aligned} P(X \geq 515) &\doteq P(\frac{X-np}{\sqrt{np(1-p)}} \geq \frac{515-508.04}{3.418}) \\ &\doteq P(Z \geq 2.036) \doteq 0.0209 \end{aligned}$$

(實際機率值約為0.0198)。在統計裡, 一件事是否夠"顯著", 乃依發生機率之大小。由於在上述假設下, 0.0209的機率算是夠小, 所以該主播之解釋並無不妥。

讀者可能注意到了, 我們所取出來討論的題目, 並未包含101年的。101年的三份考題中之機率與統計部分, 皆未包含與信賴區間有關的, 大多是排列組合或機率方面的, 這類題目較不會有爭議。而在第8題的討論中, 我們早已說"看來信賴區間快沒題目可考了"。我們的結論是, 現今大學入學考, 不論學測或指考, 機率統計的考題, 有時像在考三民主義, 思想要很制式, 才易得高分。此完全

違反由於隨機的本質, 導致各種不同的推論方法, 能百花齊放 。因此目前高中的機率統計教學, 成效實在不大 。培養出來的學生, 只怕是過於僵化, 而對隨機性的了解極少, 失去在高中數學中, 增加機率與統計題材的本意 。所以若有高中生在考試裡, 機率統計部分常無法得高分, 倒不必太沮喪, 因這並不表示沒學好 。

參考文獻

1. 黃文璋(2010). 機率論, 第二版 。華泰文化事業股份有限公司, 台北市 。
2. 黃文璋(2011). 庶民中央極限定理 。黃家小館(http://huang.nuk.edu.tw/cindex.htm) 。
3. Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K.(2011). *Probability & Statistics for Engineers and Scientists*, 9th ed. Pearson Education, Inc., Boston.