機率與統計在高中

黃文璋
國立高雄大學統計學研究所

1　前言

目前的高中數學課本, 其編寫是依據99學年度實施之"普通高級中學課程綱要"(簡稱99課綱), 這是由95學年度實施之"普通高級中學課程暫行綱要"(簡稱95暫綱)修訂而成。課綱每隔幾年會修訂一次, 此二課綱, 在教育部的網站上皆可查到。在99課綱裡, 必修分數學I、II、III, 及IV, 各4學分, 分別在高一上、下學期, 及高二上、下學期修習。選修則分標準課程、基礎課程、統整課程, 及進階課程等四類。對選修數學課程, 99課綱裡, 僅規範標準課程的綱要。自然組有數甲I, 及數甲II, 各4學分, 而社會組有數乙I, 及數乙II, 各3學分, 分別在高三上、下學期修習。這是各高中必會開設的, 因屬於指定科目考試(簡稱指考)的測驗範圍。至於其餘三類選修課程的內容, 由各校自訂。但畢竟升學至上, 既然指考不考, 則會開設的學校, 應不至於太多。

99課綱裡, 包含那些機率與統計的內容? 數學II中有四個主題, 其中有三個與機率與統計相關, 見表1。數學甲I有兩個主題, 其中第一個爲機率與統計, 見表2。另外, 數學乙I亦有兩個主題, 其中第一個仍爲機率與

統計, 見表3。假設每冊的每一主題之分量約略相等, 則機率與統計在6冊高中數學裡占了1.25冊。有多少百分比? 對自然組, 因每學期皆4學分, 故約占20.83%; 至於社會組, 因數學乙I、II各3學分, 故約占20.45%, 比重都不算低。99課綱裡中, 機率與統計的眾子題裡, 就是"抽樣與統計推論"那一子題, 其內容之一的"常態分布、信賴區間與信心水準的解讀", 幾年來一直讓高中師生有不小的困擾。

至88課綱, 皆無"常態分布、信賴區間與信心水準的解讀"。此內容是自95暫綱起, 才放進高中數學的。同時放進的, 還有"交叉分析", 這乃一完全不該在高中數學裡出現的題材, 見黃文璋(2010), 幸好在99課綱中刪除了。在99課綱的附錄裡, 對前述"解讀", 有如下説明:

> 高中程度的統計推論只做隨機變數期望值的估計, 它的背後理論是中央極限定理。要介紹中央極限定理, 就需要引入常態分布。此部分僅做通識性的介紹, 以活動方式建立學生對於中央極限定理的直觀。
>
> 對一固定的信心水準, 給出信賴區間公式, 再讓學生以亂數表模擬或實驗投擲正面出現機率爲p的銅板n次, 代入信賴區間公式, 以説明信心水準的意涵; 並以此解讀, 何以大多數的學生所得的信賴區間都會涵蓋p?

95暫綱中, 對於"解讀", 就已有類似上述的説明。雖於黃文璋(2007)一文, 我們已指出該説明不但不妥, 且可能會誤導。但令人遺憾的是, 99課綱中仍保留此説明, 只是加以修正而已。

如黃文璋(2015a)一文所指出, 對於中央極限定理, 雖不少人能琅琅上口, 但其中大多只是一知半解。事實上, 中央極限定理, 是爲了講授信賴區間而引進高中數學的。此定理高中生有能力接受嗎? 各大學碩士班甄試招生考試的面試, 短短幾分鐘能問些什麼? 試説明何謂中央極限定理, 是統計學研究所面試時, 常會出現的一道題目。參加面試的大四生, 即使大學就讀數學系或統計學系, 曾修過幾門機率或統計的課程, 對統計產生

興趣, 準備繼續深造者, 都還常將中央極限定理講得零零落落。這樣的題材, 怎會進入高中數學?

爲因應十二年國民基本教育(簡稱十二年國教), 教育部訂定“十二年國民基本教育實施計畫”, 做爲準備實施的依據。計畫中有一“提升國民素養實施方案”。根據此方案成立“國民素養專案計畫辦公室”, 並規劃出五個領域的素養, 包含語文、數學、科學、數位、教養。單獨成爲國民五個素養之一, 可看出數學的被重視。前述辦公室, 後來便提出“教育部提升國民素養專案計畫報告書”, 網路上可找到此報告書。在報告書的數學素養裡, 訂出四大數學知識素養領域, 即變化與關係、空間與形狀、數量, 及不確定性與數據。在數學四大知識素養領域中, 不確定性與數據, 便屬於機率與統計的範疇。近年來機率與統計的被重視, 可見一斑。那些機率與統計的內容, 會被放進高中數學?

早期排列組合在高中數學裡的分量很重, 考試也常出現各種刁鑽古怪的題目。表1在“排列、組合”的主題, 於備註欄中寫著“不含環狀排列”, 及“本章節要避免情境不合常理、過深、或同時涉及太多觀念的題型”。負面表列, 由此大略可看出, 太過繁瑣, 或實際上很少會有的情境之機率問題, 都不是現今高中數學裡所在乎的。取而代之, 與生活息息相關的統計, 被認爲對學生更有用。簡而言之, 制定數學課綱者認爲, 統計是現代國民該具備的知識, 因此高中時便該學習。至於太複雜的排列組合, 則不學無妨。

決定要統計後, 高中數學裡, 該涵蓋那些主題呢? 媒體上常公佈各種民調的結果, 除給出民眾對某議題之支持率外, 亦會給出抽樣誤差, 因而得到信賴區間。信賴區間看起來很重要, 於是便進入95暫綱了。做民調時會比較不同族群的支持率是否有差異, 這便是交叉分析, 也隨之進入95暫綱。而由於樣本較大時, 計算困難, 得藉助近似, 因而中央極限定理, 也就堂而皇之地引入高中數學了。要知在大學的機率與統計教科書, 中央極限定理及信賴區間, 大抵都放在全書的後半部。於經過系統的鋪

陳後, 才會進入此二題材 。如今在篇幅不多的高中數學中, 企圖介紹此概念上絕非簡單的題材, 無法講清楚, 乃屬必然 。甚至爲了探究信賴區間中, 所伴隨信心水準的涵義, 令有些好學的中學數學教師, 反而連機率的意義都弄糊塗了, 見黃文璋(2011a)及(2011b) 。

十二年國教的構想, 民國72年便提出, 經過三十餘年, 終於在民國103年正式實施 。教育部也於當年11月, 公佈"十二年國民基本教育課程綱要總綱" 。各科目據此訂定課綱, 並將自107學年度, 依照不同教育階段(國小 、國中及高中等學校一年級起), 逐年實施 。如黃文璋(2011b) 、(2011c)及(2011d)三文所指出, 目前高中數學課本中, 與信賴區間相關的內容, 可說寫得左支右絀 。因此在107課綱中, 信賴區間的刪除, 已屬勢在必行 。但鑑於統計學的重要性, 刪除這些內容後, 是否該加進什麼統計主題?

本文擬對高中數學裡的機率與統計主題, 略加探討 。首先, 信賴區間何以不適合出現在高中數學? 此點將在第2節說明 。取消信賴區間後, 若想增加新單元, 能有什麼選擇? 在第3節中, 我們將說明假設檢定(hypothesis testing)概念的重要及直觀, 適合放進高中數學 。第4節則給出一些可放進高中數學之假設檢定的例子 。最後, 我們給一簡短的結論在第5節 。

2 信賴區間在高中

統計學裡常在估計, 比較入門的是參數估計 。假設某隨機變數X遵循某一機率分佈, 分佈中有一未知的參數θ 。經由重複觀測後, 得到一組隨機樣本$X_1, X_2, \cdots, X_n$ 。一個$X_1, X_2, \cdots, X_n$的函數, 便是一統計量 。我們可以一統計量來估計θ, 這便是點估計 。由於是估計θ用, 因此統計量不可與θ有關 。其次, 以一個兩端點皆爲統計量的隨機區間來估計θ, 便是區間估計, 這樣便得到信賴區間 。至於θ會落在一信賴區間之機率, 稱爲該

信賴區間之信心水準。即使對同一分佈的同一參數, 都可有各種不同的點估計, 並未定於一尊。至於各估計量之優劣如何評比? 就依所訂定之評比標準, 標準當然也不唯一。對於信賴區間, 有時區間的端點求不出來, 或求出卻過於複雜, 不太實用。當樣本數夠大時, 常可藉助中央極限定理, 以常態分佈來近似標準化後隨機變數之和。既然是估計, 主要目的是供決策之參考, 不會太在意採取近似, 只要誤差在容忍範圍內。再度, 對同一分佈的同一參數, 信賴區間並不唯一, 通常會要求區間長度愈短愈好。

在大學的統計學課程裡, 對於信賴區間, 一般而言, 學生不會覺得太困難。因在大學裡, 於信賴區間這一章, 乃考慮各種不同的分佈, 於不同的假設(如一分佈若有兩個參數, 可分其中之一已知, 另一未知, 或兩個皆未知等情況)下, 有時要用到中央極限定理, 有時不必(前言裡所引99課綱說“背後理論是中央極限定理”, 並不正確, 當不必以常態分佈來近似, 便無涉中央極限定理), 包含的情況很多。簡單講, 學習此章時, 會感到各式各樣的分類還眞不少。若對各種情況下, 皆知如何給出一信賴區間, 這章可說便過關了。因此大學生對於信賴區間那一單元, 僅感到瑣碎, 倒不認爲難應付。

那怎麼信賴區間進入高中後, 卻吹皺一池春水, 讓原本只會讓人抱怨太難的高中數學, 變成有一不算難, 但卻比難還令人更受不得的單元? 我們猜想主要原因如下。

高中的信賴區間, 只針對二項分佈。二項分佈是一簡單, 但機率值之和不好計算的分佈。因此高中數學從95暫綱起, 便先講中央極限定理, 以常態分佈來近似二項分佈。這是高中引進中央極限定理的唯一理由, 且只考慮幾乎可說是最簡單的版本, 即只適用二項分佈。由於其證明超過高中範圍, 教科書遂藉助圖示, 以使學生了解此定理到底在說些什麼。假設S_n有參數n, p的二項分佈, 即$\mathcal{B}(n,p)$分佈。有些書中繪出某一S_n的直方圖, 然後說, n夠大時, 會近似標準常態分佈的圖形。這當然是錯的, 要

將S_n標準化才行, 否則當n愈大, S_n的直方圖會愈像一水平線。事實上, 比前述講法更離譜的還有不少, 見黃文璋(2011d)。大部分教科書的作者可能是想當然耳, 並未眞正讓n逐漸增大, 以觀察S_n直方圖的變化。這不表示在大學學習中央極限定理時, 就一路平坦, 通行無阻。而是不像在高中的教科書中, 就那麼一個孤零零的定理, 大學裡有很多題材要講。因此通常大學教師並不太在乎什麼圖示, 快速便帶過中央極限定理了, 最多藉助特徵函數(characteristic function)來證明。也就是學習方面潛藏的問題, 在大學裡較不會浮現。

高中數學, 於中央極限定理之後, 便是應用。再度, 只有一個應用, 就是做民調時求信賴區間。但民調裡的抽樣, 通常是所謂取出後不放回, 如此各樣本不獨立。因而涉及的分佈, 爲超幾何分佈(hypergeometric distribution), 而非二項分佈, 如此中央極限定理便不適用了。這部分可把教師弄糊塗了。數學上何曾有在條件不滿足之下, 定理仍可引用? 由於是在高中的“數學”課程中, 少有教科書的作者, 能理直氣壯地將此處用到的近似概念講清楚。事實上, 是有些認眞的教師, 試圖補上自認該有的“證明”, 只是其證明當然都是錯的, 見黃文璋(2011d)一文的第4節。欲將書上含糊處弄清, 往往引出更多問題, 導致教師無所適從。

高中數學裡求信賴區間, 就是只局限在$\mathcal{B}(n,p)$分佈(而且如前所述, 其實是超幾何分佈)中的p, 且信心水準就設定爲95%, 沒其他變化, 很單純。不像大學的統計學, 在信賴區間那一章, 情況多到會讓人眼花撩亂。高中裡涵蓋的內容不多, 一下子講完事小, 變化那麼少, 怎麼考試? 而且還考選擇題? 這是大學統計學教師不會有的煩惱, 因種類夠多, 且考題多半是計算的型式。在黃文璋(2011e)一文中, 我們將民國98至101年, 那4年的學測數學, 及指考數學甲與數學乙, 共12份試題中, 值得商榷的機率與統計的考題, 皆取出來分析。數學國文化, 可看到有關信賴區間的考題, 差不多都成爲文字遊戲了。所以, 若聽到有大學的統計學教授說, 某次指考那道信賴區間的考題, 5個選項他錯了3個, 也就不必感到太訝異

了。因國文科的考題, 統計學教授考不好, 乃稀鬆平常。只是這樣子學習數學, 有何樂趣呢?

至於信心水準, 所造成的困惑, 也絕對不會少。長期以來, 高中數學裡出現的機率, 大抵是古典機率, 即基於相同的可能性。因此對機率的理解, 可說就是排列組合加上除法。連條件機率, 恐怕也只會讓人連想到除法。因此往昔在高中數學裡接觸機率時, 師生大約皆很少會去思索機率的意義是什麼? 引進信賴區間後, 一個信心水準95%的信賴區間, 可能不包含欲估計的參數。那95%是指什麼? 而且怎麼每次得到的信賴區間都不同? 到底在信賴什麼? 師生皆一頭霧水, 心想, 這是什麼數學? 又從信賴區間起, 有了取樣, 之後不稱機率, 卻稱信心水準。這又產生一些效應。抽屜裡有一支鉛筆跟一支鋼筆, 某師拿出鉛筆放背後, 然後問學生, 他背後是鉛筆及鋼筆的機率, 各爲若干。學了信賴區間後, 有些教師連此題是否有意義, 都感到極度困惑。歸根結柢, 長期以來, 在高中數學中, 不曾認眞說明機率的涵義, 條件機率更是模糊以對, 從未強調機率值會變的概念。甚至, 機率與統計中, 最重要的隨機性, 也常被忽視。在前言裡所引99課綱說明的第二段中, 首先"給出信賴區間公式", 這種說法就未傳遞統計學裡, 可有南轅北轍, 各種不同的估計法, 而由於隨機性, 使各估計法常互有較準的時候; 其次, "以此解讀, 何以大多數的學生所得的信賴區間都會涵蓋p? "更是連隨機性都不顧了。這樣子學統計, 眞是教壞學生。

數學家習於講數學之美, 數學裡強調準確性。但對高中數學教師, 於信賴區間這一單元, 絲毫體會不到美感, 準確性也不知何在? 數學課程裡爲何要有此內容? 不少教師無法說服自己, 因此在很多場合, 屢有教師建議取消此主題。怎會只因常在媒體上看到民調結果的公佈, 就覺得信賴區間很重要, 高中生該會? 當初將信賴區間引進高中數學, 可說未經深思。對大部分的國民, 那有機會去執行民調, 去抽取1千多個樣本? 信賴區間其實不必成爲國民的基本數學素養。何況, 既然這麼多年下來, 一直

講不清楚, 且連教師都弄不太懂的內容, 講授又何益? 是該自高中數學裡刪除了。

3 假設檢定

信賴區間自高中數學移出後, 是否須加進什麼主題? 什麼都不加自然是一個選項。這選項其實並不壞, 因學習最怕的是被教壞, 而非學太少。要知統計學的內涵並不易吸收, 連很多基本概念都有點深。即使在大學裡, 一般認爲統計學比微積分還難教。因諸如微分及積分等, 都有物理意義, 只要藉助圖形, 很容易便能說明。但在統計學中, 連期望值都很難解釋清楚。所以統計學家, 對於將機率與統計引進高中, 多半興趣缺缺。但有些熱心中小學數學教育的數學家, 常覺得“不確定性”很重要, 以爲中小學生該多學些機率與統計。因此一旦拿掉信賴區間, 會爲高中數學產生一缺口, 因而忐忑不安。若眞得補個主題, 那就是假設檢定了。

對統計學略有概念者, 獲知這樣的建議, 可能會感到驚訝。因在大學的統計學課程裡, 學生視假設檢定比信賴區間難多了。至於過去排斥信賴區間的高中數學教師, 曉得這是刪除信賴區間後的代價, 說不定將改變主意了。因好不容易才弄懂信賴區間, 怎麼要換個更難的主題進來? 眞是前門拒虎, 後門進狼。

首先, 企圖在高中數學有限的篇幅中, 將信賴區間及相關的題材都交待清楚, 幾乎是不可能的任務。從一開始的中央極限定理, 即使學養豐富的高中數學教師, 對課本中的圖示法, 便常百思不解。以高度表示機率, 屬於離散型之二項分佈的直方圖, 如何有辦法趨近至以面積表示機率, 屬於連續型之常態分佈的機率密度函數之圖形? 實在沒有必要讓高中師生陷入這種困境。而且, 就算眞弄懂圖示法後又如何? 英雄無用武之地, 對了解後續的信賴區間, 乃毫無幫助。要知高中數學裡, 連機率的涵義, 都一向未曾細究, 那信心水準之意義, 怎可能講明白? 那些自以爲已能掌握

信賴區間之內涵的中學教師, 恐怕有極高的比例, 只是教熟而已。既然教師難以講解、學生難以吸收, 且日後用到的機會不大, 整個高中階段信賴區間的學習, 可說完全是浪費。在高中數學裡, 這是一連雞肋都稱不上的主題, 棄之不必惋惜。

是有一些原因支持我們將假設檢定放進高中數學。首先, 其概念遠比信賴區間容易, 也更有用。但最重要的是, 假設檢定原本就是人們做決策時常採的一種思維, 相當直觀。事實上, 如果去翻閱大學統計學的教科書, 將發現不少是將信賴區間的章節, 置於假設檢定之後。因學完假設檢定後, 信賴區間便水到渠成了。一旦高中數學引進假設檢定, 尚可用來檢驗機率值是否合理。如此機率從上游的定義, 到下游的檢定, 便有一套完整的體系。學生將更易於了解機率的意義, 進而理解隨機性的內涵。能弄清楚機率與隨機性的意義, 高中機率與統計的教學, 便算成功了。

設某款汽車, 汽油每1公升可行駛14公里, 問20公升可行駛幾公里? 答案當然是280公里。你不必多想車子是否能如此省油, 且能1公升14公里維持不變。這是數學問題, 對於所給的假設, 接受便是, 不必有其他想法。但若有某汽車商, 宣稱該公司的某款汽車, 汽油每1公升平均可行駛14公里, 則說不定會被消費者檢舉廣告不實, 因他們所買的該款車, 耗油多了。

數學裡常給假設, 在某些假設下, 做些推導或證明。在眞實的人生裡, 可難以靠著假設走天下。數學裡可假設42取6的樂透彩, 頭獎號碼爲隨機產生, 問開出的頭獎號碼有連號之機率。對彩券迷, 卻有不少就是不相信中獎全憑機運, 他們下注時, 一向精心地挑選號碼。眞的沒有明牌嗎? 令人好奇。很多人想瘦身, 琳琅滿目的方法, 不知何者可信? 某減肥藥的代理商, 找來很多瘦身成功者現身說法, 個個服藥後, 兩星期內就至少減5公斤。效果眞有那麼好? 雖然心動, 難免仍半信半疑。鯨類中最大的座頭鯨(humpback whale), 由於濫捕, 數量逐漸降低, 曾瀕臨絕種。Barlow, *et al.* (2011)一文宣稱, 近年來因保育成功, 數量已慢慢回升。座頭鯨生

活在大海裡, 既無法全抓起來數, 如何能確定增加?

很多時候, 就是難以證明所述爲眞或爲僞。法庭裡也一樣, 被起訴者究竟是清白或有罪? 檢察官及辯護律師, 皆掌握一些對己方有利的證據。雖號稱我心如秤, 由判決最後常只好依法官投票決定, 就知連法官的意見都不見得一致。數學中處處是證明, 但在各行各業裡, 經常面對的是無法判定眞僞之情境。眞相如何, 往往只有天曉得。但上天比法官更不語, 人們只能自行決定要接受那一選擇。這種眞僞難辨的情況, 日常生活中亦屢遭遇。例如, 早上出門需要帶傘嗎? 氣象局的天氣預報裡, 除溫度外, 亦有降雨機率。機率才10%, 傘可以不帶吧! 等等, 過去公佈降雨機率10%的日子, 似乎常下雨, 讓人淋了一身雨。氣象局的10%, 是否都低估? 令人存疑。但會不會是人們對沒帶傘卻下雨, 印象比較深? 氣象局的預測, 還是可信。有人信有人不信, 只是可能永遠無法證明10%的機率, 是否正確。

處在此隨機的世界, 很多事都說不得準。一切都是假設, 端看你接受那一個? 如何做出決定? 人們常說不可先射箭再畫靶, 先入爲主總是無法讓人心服。所以如何接受, 宜比照今日法官判案的原則, 即採無罪推定。被起訴之涉嫌收賄官員, 明明千夫所指, 法官卻會先假設他無罪、是無辜的。在無罪的假設下, 何以一再配合廠商修改規格? 銀行户頭何以有多筆巨款匯入? 這很不尋常, 被起訴者得好好說明。而舉證之所在, 敗訴之所在。若解釋不清, 被判有罪就怨不得人了。

不尋常的事件, 或說顯著事件, 乃指發生機率很小的事件。顯著事件若發生, 會引人注意。尋常的事件, 即發生機率不算低的事件, 其發生自然不會令人太在意。譬如說, 一早發現系辦公室前夜失竊, 查看監視器, 發現某生半夜1點走往系辦公室, 這很不尋常, 該找此生來問問。假設檢定的想法, 就是在無罪推定的原則下, 如果某顯著事件發生, 那原本的假設, 就可放棄了。至於多小的機率才算顯著? 乃視情況而定。若是食品添加物超過的含量, 5%的發生機率都不算小, 因總要給廠商警惕; 若是死

刑的判決, 連0.01%誤判的機率都覺得太大, 畢竟人命關天。

法官會不會誤判? 當然會, 包青天只存在小說或戲劇裡。會有什麼樣的誤判? 有兩類。第一類是無罪卻被判有罪, 第二類是有罪卻被判無罪。民主時代, 相當注重人權, 通常第一類誤判被視爲較嚴重。雖兩類誤判都不應該發生, 都須儘量減少, 但除非採取增加樣本數等措施, 否則常很難將兩類誤判的機率同時降低。想想如果過度把關第一類誤判, 則實際有罪, 卻由於證據不夠強, 因而被縱放的, 將大幅度地增加, 這並非好事; 反之, 若過度把關第二類誤判, 則將導致寧可錯殺一千, 不可放過一人的後果, 製造出無數冤屈。

統計學裡, 便依以上所述, 無罪推定之客觀思維, 發展出一套假設檢定的架構。在此架構中, 有兩個假設, 其中一個是虛無假設(null hypothesis), 以H_0表之, 另一個是對立假設(alternative hypothesis), 以H_a表之。前者通常表現況, 或我們傾向不相信的; 後者則通常表我們傾向相信的。依所取的樣本, 來決定接受虛無假設或對立假設。虛無假設爲眞, 卻接受對立假設, 稱爲第一型錯誤; 對立假設爲眞, 卻接受虛無假設, 稱爲第二型錯誤。先設定一能容忍的第一型錯誤之機率, 以α表之。常取的α值爲0.01, 0.05, 或0.1等。給定α後, 便要決定拒絕域, 即決定何時拒絕虛無假設而接受對立假設。拒絕域須使第一型錯誤的機率不超過α, 但愈接近α愈好, 因拒絕域愈大接受域便愈小, 因而第二型錯誤的機率也就愈小。就好像拿到相同的數據, 可有不同的推論, 對相同的α, 不同的人可給出不同的拒絕域。在同一α下, 最理想的狀況是, 找到使第二型錯誤的機率最小之拒絕域, 此稱爲最佳拒絕域。那些以爲假設檢定的主題很難弄懂者, 往往是對找最佳拒絕域心有餘悸。不過在很多情況下, 憑直觀所決定的拒絕域, 便是最佳拒絕域。至於虛無假設命名的由來, 乃是因那根本是一空的假設。試想如果產品明明合格, 主管的政府單位, 卻懷疑其成分有問題, 非要抽樣來檢驗, 最後卻宣佈該產品合格, 廠商不罵擾民、損害商譽才怪。接受虛無假設, 表示白忙一場, 天下本無

事, 庸人自擾之。虛無假設是執行檢定者, 一點都不想接受的假設。

科學家不時宣佈一些新發現。如接觸殺蟲劑會使人罹患帕金森氏症的機率增加, 過胖的中年人罹患失智症的風險較低等。到底增加或減少, 憑藉的依據, 可說就是執行一項假設檢定後的推論。人們經常在做決策, 決定該接受那一選項? 往往是依假設檢定的思維。假設檢定是一適合讓國民提早學習的統計方法。想稍進一步了解其內容的讀者, 可參考黃文璋(2004)及(2005)二文。

4　高中數學假設檢定之例

假設檢定涵蓋的範圍相當廣泛, 還有專書整本都講假設檢定。在高中數學裡, 雖只宜給些基本的介紹, 但爲利於教學, 仍得要有夠多的變化。主要用來檢定機率值, 且主要涉及兩個相當簡單的分佈, 即二項分佈與幾何分佈, 應是一合理的安排。

人們常在談機率, 機率到底是什麼? 每個人心中所想並不盡相同, 但大致有下述幾種不同的意義。先看第一種。骰子有6個面, 何以常就將每個面出現的機率都當成$1/6$? 原因乃假設骰子爲公正, 亦即每個面出現的機率都相同, 於是就各$1/6$。一副撲克牌有52張, 玩梭哈遊戲, 每人各發5張, 求某君拿到4張A的機率。條件不足, 怎麼算呢? 其實並沒那麼複雜。假設撲克牌洗均勻了, 則每一種組合出現的機率便可假設皆相同。而共有$\binom{52}{5}$種組合, 所以每一種組合出現的機率者都是$1/\binom{52}{5}$。又5張牌中有4張A, 共有48種可能, 故所求爲$48/\binom{52}{5}$。在前述這類例子中, 都是以相同的可能性來解釋機率。先看共有幾種可能出現的結果, 當然得有限才行, 且每種出現的機率皆假設相同。這種機率模式, 又稱古典模式, 在第2節已提過。雖說古典, 當沒有其他資訊時, 至今仍常採用。譬如說, 某君參加一項面試, 5位只錄取1位, 他環顧一下, 覺得每人都差不多, 遂認爲自己錄取的機率爲$1/5$。其次看第二種意義。某

支職棒隊有位打擊好手, 記錄顯示其打擊率爲0.352。在一關鍵時刻, 輪到他上場打擊, 只要揮出安打球隊便贏了。他獲安打的機率爲何? 不少人以爲是0.352。但也有人很樂觀, 賭該選手一定擊出安打。要知機率的意義, 並無法由少數幾次的結果來評斷。若是長期觀看球賽, 便會體會0.352爲一合理的答案。這便是頻率的解釋, 適用於可重覆觀測的事件, 以發生次數除以觀測次數, 也就是事件發生次數的相對頻率, 來代表機率。諸如有些專家所給出婦女懷孕生雙胞胎的機率等, 大抵是源自於頻率之想法。再來看第三種意義。某男生想追一心儀的女孩, 先評估追上的機率, 認爲有0.7。這0.7的機率如何產生? 雖有衡量雙方的條件, 及兩人平常的互動情形, 但顯然是相當主觀的, 這便是對機率主觀的解釋。相較於主觀, 頻率的解釋, 算是客觀多了, 所以又稱客觀的解釋。

以上三種, 大致是長期以來, 人們對機率的解釋。二十世紀以後, 如同以公理化的方式定義實數系統, 一套經由公理化的方式, 以定義機率的模式產生了, 詳情可查閱一般機率論的書。

以p表某一銅板投擲後出現正面之機率。可以因銅板有兩個面, 而以爲$p = 0.5$; 也可因投擲多次後, 發現正面出現的相對頻率爲0.5, 而得$p = 0.5$; 也可主觀上就認定$p = 0.5$。對一事件, 不論以那種方式定義其發生的機率, 只要是如銅板的投擲, 可以重覆觀測, 便能對發生機率, 做一假設檢定。底下就以銅板爲例。

有人懷疑p不是0.5, 而可能是0.6, 想做一檢定。遂取$H_0 : p = 0.5$, $H_a : p = 0.6$。隨機投擲銅板6次, 令X表出現正面的次數。因p愈大, 愈容易出現正面, 所以直觀上X較大時, 會拒絕H_0, 而接受H_a。故拒絕域常會取成$\{X \geq c\}$的形式, 其中$c = 0, 1, 2, \cdots, 6, 7$。

例1. 取拒絕域爲$\{X \geq 4\}$。分別求兩型錯誤的機率。

解. 當$p = 0.5$, X有$\mathcal{B}(6, 0.5)$分佈。當$p = 0.6$, X有$\mathcal{B}(6, 0.6)$分佈。故第一型錯誤的機率爲

$$P(X \geq 4|p=0.5) \;=\; \binom{6}{4}0.5^6 + \binom{6}{5}0.5^6 + \binom{6}{6}0.5^6 = 0.34375\text{ 。}$$

至於第二型錯誤的機率爲

$$\begin{aligned} P(X \leq 3|p=0.6) \;=\; & \binom{6}{0}0.4^6 + \binom{6}{1}0.6^1 \cdot 0.4^5 + \binom{6}{2}0.6^2 \cdot 0.4^4 \\ & + \binom{6}{3}0.6^3 \cdot 0.4^3 = 0.45568\text{ 。} \end{aligned}$$

在上例中, 可看出第一型錯誤的機率超過0.3, 不算太小, 第二型錯誤的機率則更大, 超過0.4 。下例取較小的拒絕域, 以爲比較 。

例2. 取拒絕域爲$\{X \geq 5\}$, 重做上題 。
解. 第一型錯誤的機率爲

$$P(X \geq 5|p=0.5) = \binom{6}{5}0.5^6 + \binom{6}{6}0.5^6 = 0.109375\text{ 。}$$

至於第二型錯誤的機率爲

$$\begin{aligned} P(X \leq 4|p=0.6) \;=\; & \binom{6}{0}0.4^6 + \binom{6}{1}0.6^1 \cdot 0.4^5 \\ & + \binom{6}{2}0.6^2 \cdot 0.4^4 + \binom{6}{3}0.6^3 \cdot 0.4^3 \\ & + \binom{6}{4}0.6^4 \cdot 0.4^2 = 0.76672\text{ 。} \end{aligned}$$

在例2中, 雖第一型錯誤的機率較例1中的降低, 但第二型錯誤的機率卻更大了 。在保護虛無假設, 壓低第一型錯誤的機率之情況下, 有時會犧牲第二型錯誤的機率 。也就是說, 對立假設爲眞, 卻接受虛無假設的機率將升高 。以法庭判決來說明, 爲了儘量避免冤屈, 便難免有不小比率的被起訴者, 是明明有罪(對立假設爲眞), 卻被法官宣判無罪(接受虛無假設) 。

例3. 取拒絕域爲$\{X \le 4\}$, 重做上例。
解. 第一型錯誤的機率爲

$$\begin{aligned} P(X \le 4|p=0.5) \;=\; & \binom{6}{0}0.5^6+\binom{6}{1}0.5^6+\binom{6}{2}0.5^6 \\ & +\binom{6}{3}0.5^6+\binom{6}{4}0.5^6=0.890625\text{。} \end{aligned}$$

至於第二型錯誤的機率爲

$$P(X \ge 5|p=0.6) \;=\; \binom{6}{5}0.6^5\cdot 0.4^1+\binom{6}{6}0.6^6=0.23328\text{。}$$

在上例中, X值較小時反而拒絕$p=0.5$, 接受$p=0.6$, 這樣的拒絕域並不合理, 因此導致第一型錯誤的機率高達0.89多。但第二型錯誤的機率卻降低了。又由例2知, $\{X \ge 5\}$爲給定$\alpha=0.11$下之一拒絕域。但對$\alpha=0.11$, 尚有其他拒絕域, 見下例。

例4. (1)試證拒絕域若取爲$\{X=6\}$, $\{X=0\}$, 或$\{X \le 1\}$, 則第一型錯誤的機率皆不超過0.11。(2)對前述3個, 及$\{X \ge 5\}$等4個第一型錯誤的機率不超過0.11之拒絕域, 求出第二型錯誤的機率最小者。
解(1).

$$P(X=6|p=0.5)=\binom{6}{6}0.5^6=0.015625<0.11\text{。}$$
$$P(X=0|p=0.5)=\binom{6}{0}0.5^6=0.015625<0.11\text{。}$$
$$P(X\le 1|p=0.5)=\binom{6}{0}0.5^6+\binom{6}{1}0.5^6=0.109375<0.11\text{。}$$

故3個拒絕域皆使第一型錯誤的機率不超過0.11。
(2). 首先, 在例2中已求出

$$P(X \le 4|p=0.6)=0.76672\text{。}$$

其次,

$$
\begin{aligned}
P(X \leq 5|p=0.6) &= 1 - P(X=6|p=0.6) \\
&= 1 - \binom{6}{6} 0.6^6 = 0.953344 \text{ 。}
\end{aligned}
$$

又

$$
\begin{aligned}
P(X \geq 1|p=0.6) &= 1 - P(X=0|p=0.6) \\
&= 1 - \binom{6}{0} 0.4^6 = 0.995904 \text{ 。}
\end{aligned}
$$

最後,

$$
\begin{aligned}
P(X \geq 2|p=0.6) &= 1 - P(X=0|p=0.6) - P(X=1|p=0.6) \\
&= 1 - \binom{6}{0} 0.4^6 - \binom{6}{1} 0.6^1 \cdot 0.4^5 = 0.95904 \text{ 。}
\end{aligned}
$$

故4個拒絕域中, 使第二型錯誤的機率最小者爲$\{X \geq 5\}$ 。

上例的結果並不難猜測到 。首先, 如前所述, 對於此處之H_0及H_a, 直觀上, 合理的拒絕域有$\{X \geq c\}$的形式, 其中$c = 0, 1, 2, \cdots, 6, 7$ 。當c愈大, 表拒絕域愈小, 如此第一型錯誤的機率愈小, 且第二型錯誤的機率愈大 。若$c = 0$, 表必拒絕H_0, 因此第一型錯誤的機率爲1, 且第二型錯誤的機率爲0; 至於$c = 7$, 表拒絕域爲空集合, 即永不拒絕H_0, 因此第一型錯誤的機率爲0, 且第二型錯誤的機率爲1 。對其餘的c值, 兩型錯誤的機率皆介於0與1之間 。若取$\{X \geq c\}$形式的拒絕域, 則對一給定的α, 存在一c, 使第一型錯誤的機率, 最接近但不超過α 。對拒絕域最大者, 其接受域顯然最小, 故此時第二型錯誤的機率, 爲在α下最小 。即得α下最佳的$\{X \geq c\}$形式之拒絕域 。事實上, 這就是α下的最佳拒絕域, 不過此處不證明 。

例5. 承上例 。試分別對(1)$\alpha = 0.05$, (2)$\alpha = 0.11$, (3)$\alpha = 0.2$, 在$\{X \geq c\}$形式的拒絕域中, 找出最佳者 。

解. 當$c=4$時, 第一型錯誤的機率爲

$$\begin{aligned} P(X \geq 4|p=0.5) &= \binom{6}{4}0.5^6 + \binom{6}{5}0.5^6 + \binom{6}{6}0.5^6 \\ &= 0.34375 > 0.2 \text{ 。} \end{aligned}$$

當$c=5$時, 例2中已給出第一型錯誤的機率爲$0.109375 > 0.05$。當$c=6$時, 例4中已給出第一型錯誤的機率爲$0.015625 \leq 0.05$。故對(1), $\{X \geq c\}$形式的拒絕域中, 最佳者爲$\{X=6\}$。對(2)及(3)的α, $\{X \geq c\}$形式的拒絕域中, 最佳者皆爲$\{X \geq 5\}$。

對8個$\{X \geq c\}$形式的拒絕域, 我們可按α分別落在的區間, 找出最佳拒絕域。此部分留給讀者自行完成。另外, 亦可有諸如$\{X=1,3,5\}$的拒絕域。可以想見, 這種不會是太好的拒絕域。

例6. 設$\alpha=0.11$, 且取$\{X \geq c\}$形式的最佳拒絕域。試分別對觀測到$X=4$, $X=5$, $X=6$, 決定該接受H_0或H_a?
解. 由例5, 當$\alpha=0.11$時, $\{X \geq c\}$形式的最佳拒絕域爲$\{X \geq 5\}$。若$X=4$, 因並未落在拒絕域, 故接受H_0。至於$X=5$, 與$X=6$, 皆落在拒絕域, 故皆拒絕H_0且接受H_a。

在上述幾個例子中, 由於只是展示用, 不擬讓計算過於繁瑣, 故銅板之投擲次數均取爲少少的6次。實際執行檢定時, 投擲次數通常會取得大些, 一方面降低第二型錯誤的機率, 一方面可有較多不同的第一型錯誤之機率的拒絕域。

我們亦可藉助幾何分佈來執行檢定, 特別是當p較小時。取$H_0: p=0.01$, $H_a: p=0.03$。隨機投擲銅枚, 直至出現1個正面便停止。令Y表總共投擲的次數。直觀上Y較小時會拒絕H_0, 而接受H_a。Y有$\mathcal{G}e(p)$分佈, 即參數p之幾何分佈, 且

$$P(Y=k) = p(1-p)^{k-1}, k=1,2,\cdots \text{ 。}$$

又

$$
\begin{aligned}
&P(Y \le k) = 1-(1-p)^k, k = 1, 2, \cdots,\\
&P(Y > k) = (1-p)^k, k = 1, 2, \cdots,\\
&P(Y \ge k) = (1-p)^{k-1}, k = 1, 2, \cdots,\\
&P(Y < k) = 1-(1-p)^{k-1}, k = 1, 2, \cdots 。
\end{aligned}
$$

例7. 取拒絕域爲$\{Y \le k\}$。對$\alpha = 0.05$, 求出最大的整數k, 並分別求此時兩型錯誤的機率。

解. 第一型錯誤的機率須滿足

$$P(Y \le k|p = 0.01) = 1-(1-0.01)^k = 1-0.99^k \le 0.05 。$$

解出$k \le 5.10\cdots$。故最大的整數$k = 5$。此時第一型錯誤的機率爲

$$1-0.99^5 \approx 0.0490 。$$

至於第二型錯誤的機率爲

$$P(Y > 5|p = 0.03) = (1-0.03)^5 = 0.97^5 \approx 0.8587 。$$

所以, 若取$\alpha = 0.05$, 及拒絕域$\{Y \le 5\}$, 則當觀測到$Y = 4$, 便拒絕H_0, 且接受H_a; 若得$Y = 6$, 便接受H_0。另外, 亦可反過來, 考慮$H_0 : p = 0.03$, $H_a : p = 0.01$。直觀上, 此時Y較大時該拒絕H_0, 而接受H_a。見下例。

例8. 取拒絕域爲$\{Y \ge k\}$。對$\alpha = 0.05$, 求出最小的整數k, 並分別求此時兩型錯誤的機率。

解. 第一型錯誤的機率須滿足

$$P(Y \ge k|p = 0.03) = (1-0.03)^{k-1} = 0.97^{k-1} \le 0.05 。$$

解此$k \geq 99.35\cdots$。故最小的整數$k=100$, 即拒絕域爲$\{Y \geq 100\}$。此時第一型錯誤的機率爲

$$0.97^{99} \approx 0.0490 < 0.05 \text{。}$$

至於第二型錯誤的機率爲

$$P(Y < 100|p = 0.01) = 1 - (1 - 0.01)^{99} = 1 - 0.99^{99} \approx 0.6303 \text{。}$$

最後來看, 亦可以二項分佈來檢定例7中的p。取$H_0 : p = 0.01$, $H_a : p = 0.03$。再度隨機投擲銅板6次, 令X表出現正面的次數, 且取$\{X \geq c\}$形式的拒絕域。因

$$P(X \geq 1|p = 0.01) = 1 - P(X = 0|p = 0.01) = 1 - 0.99^6 \approx 0.0585,$$

故只要$0.0586 < \alpha < 1$, 皆可取拒絕域爲$\{X \geq 1\}$。此時第二型錯誤的機率爲

$$P(X = 0|p = 0.03) = 0.97^6 \approx 0.8330 \text{。}$$

又因

$$\begin{aligned} P(X \geq 2|p = 0.01) &= 1 - P(X = 0|p = 0.01) - P(X = 1|p = 0.01) \\ &= 1 - 0.99^6 - 6 \cdot 0.01^1 \cdot 0.99^5 \approx 0.00146, \end{aligned}$$

故若$\alpha = 0.05$, 則可取拒絕域爲$\{X \geq 2\}$。此時第二型錯誤的機率爲

$$\begin{aligned} P(X \leq 1|p = 0.03) &= P(X = 0|p = 0.03) + P(X = 1|p = 0.03) \\ &= 0.97^6 + 6 \cdot 0.03^1 \cdot 0.97^5 \approx 0.9875 \text{。} \end{aligned}$$

第二型錯誤的機率可說相當大。

至於以二項分佈來檢定例8中的p, 就留給讀者自行完成。

5　結論

統計學的理論豐富, 極具應用價值。只是有些看似簡單的統計方法或概念, 於實際應用時, 卻會出現很多料想不到的問題。進行民調以得信賴區間便是一例。從一有紅球及白球的袋子中, 經由簡單隨機抽樣, 以估計紅球所占比率, 並得信賴區間, 感覺上較無問題。不過還是得確保袋中的球, 果眞已攪拌均勻, 且每次能做到"隨機"取球。至於若要經由簡單隨機抽樣, 以調查某地區民眾對某議題的支持率, 則將發現調查人與球, 乃完全是兩回事。人會找不到、會拒訪、會不誠實回答等。

即使看來有明確定義的百分位數, 也經不起細究。原本是針對幾乎可視爲連續的大量數據, 而且因只是爲了對數據的散佈, 能掌握一些粗略地概念, 故近似是允許的。引進中學後, 爲便於學生計算, 所處理的往往是少量數據。而且中學數學裡, 也不易說明諸如差不多, 及大致等的概念, 何時能用。因此依現有定義, 連1, 2, 3, 4, 5這組數據的中位數都不存在, 完全違反一般人的直觀。有關百分位數, 所引起的問題還不少, 可參黃文璋(2015b)一文。

高中數學現有的數據分析那一主題, 教學上並無太大爭議, 所以保留無妨。只是涉及數據處理時, 仍須謹慎爲之。像104學年的學測數學, 便有一不太恰當的考題, 見黃文璋(2015c)一文。

由於近年高中機率與統計的教學, 不時產生種種大小爭議, 讓若干統計學者以爲, 時機就是未到, 現今在高中數學中, 尚不適合放進太多機率與統計的內容。甚至以爲, 若能把條件機率講清楚, 讓學生了解機率值有可能會隨著新增資訊而變, 教學就已算有相當成效了。要知在隨機世界中, 眞相本難知, 一切都是假設, 看你接受那一個。因此對機率值的認定有所改變, 並不必感到太奇怪。此正如當有新事證出現, 法庭也可能改變判決。這是統計學與數學之一大差別。數學裡重視不變, 由給定的假設, 依一定的邏輯, 推導出結果。一旦推導出來, 就毫無例外的成立。而對所

給的假設, 接受便是, 不必多質疑。至於統計學, 則是由所獲得的結果, 倒過來, 檢驗當初的假設, 是否可接受。假設檢定設計出之那套接受或拒絕的程序, 符合人們向來客觀做決策時的思維。也是一種較科學、不先入爲主、儘量爲對方設想之思維。因此, 若眞要在高中數學裡, 加進學生日後較有用, 且高中階段較易接受的統計內容, 則假設檢定便爲一適當的主題。思維秉持無罪推定的原則, 也算是國民宜具備的基本素養之一。況且, 引進假設檢定後, 可檢定機率值, 將有助於讓學生更了解機率的涵義。

表1 數學II(有限數學)、4學分

主題	子題	內容	備註
二、排列、組合	1. 邏輯、集合與計數原理	1.1 簡單的邏輯概念: 介紹「或」、「且」、「否定」及笛摩根定律 1.2 集合的定義、集合的表示法與操作 1.3 基本計數原理(含窮舉法、樹狀圖、一一對應原理) 1.4 加法原理、乘法原理、取捨原理	
	2. 排列與組合	2.1 直線排列、重複排列 2.2 組合、重複組合	2.1 不含環狀排列 本章節要避免情境不合常理、過深、或同時涉及太多觀念的題型
	3. 二項式定理	3.1 以組合概念導出二項式定理、巴斯卡三角形	3.1 不含超過二項的展開式

表1 數學II(有限數學)、4學分(續)

主題	子題	內容	備註
三、機率	1. 樣本空間與事件 2. 機率的定義與性質 3. 條件機率與貝氏定理	1.1 樣本空間與事件 2.1 古典機率的定義與性質 3.1 條件機率、貝氏定理、獨立事件	 2.1 不含幾何機率
四、數據分析	1. 一維數據分析 2. 二維數據分析	1.1 平均數、標準差、數據標準化 2.1 散佈圖、相關係數、最小平方法	1.1 只談母體數據分析，不涉及抽樣，可用計算工具操作 2.1 可用計算工具操作。最小平方法的證明置於附錄
附錄	2. 最小平方法	最小平方法的證明	

表2 數學甲I(4學分)

主題	子題	內容	備註
一、機率統計	1. 隨機的意義	1.1 隨機的意義	
		1.2 期望值、變異數、標準差	
	2. 二項分布	2.1 獨立事件、重複試驗、二項分布、二項分布的性質	
	3. 抽樣與統計推論	3.1 抽樣方法：簡單隨機抽樣	3.1 不含系統抽樣、部落抽樣
		3.2 亂數表	
		3.3 常態分布、信賴區間與信心水準的解讀	

表3 數學乙I(3學分)

主題	子題	內容	備註
一、機率統計	1. 隨機的意義	1.1 隨機的意義	
	2. 期望值、變異數、標準差	2.1 期望值、變異數、標準差	
	3. 獨立事件	3.1 獨立事件	
	4. 二項分布	4.1 重複試驗、二項分布、二項分布的性質	
	5. 抽樣與統計推論	5.1 抽樣方法: 簡單隨機抽樣	5.1 不含系統抽樣、部落抽樣
		5.2 亂數表	
		5.3 常態分布、信賴區間與信心水準的解讀	

參考文獻

1. 黃文璋(2004). 統計學裡無罪推定的精神. 科學發展月刊, 383期(2004年11月號): 68-73.
2. 黃文璋(2005). 統計顯著性. 數學傳播季刊, 29(4): 29-38.
3. 黃文璋(2007). 統計裡的信賴. 數學傳播季刊, 30(4): 48-61.

4. 黃文璋(2010). 談高中數學裡的交叉分析. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm).
5. 黃文璋(2011a). 對機率要有信心. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm).
6. 黃文璋(2011b). 認識機率. 數學傳播季刊, 35(2): 32-44.
7. 黃文璋(2011c). 關於中央極限定理之圖示. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm).
8. 黃文璋(2011d). 庶民中央極限定理. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm).
9. 黃文璋(2011e). 機率統計考題探討. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm)—各式文章—統計集注.
10. 黃文璋(2015a). 以中央極限定理選才? 科學人. 156期(2015年2月號): 24.
11. 黃文璋(2015b). 數據素養. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm).
12. 黃文璋(2015c). 104年學測數學科的一道統計題. 黃家小館(http://www.stat.nuk.edu.tw/huangwj/cindex.htm).
13. Barlow, J., Calambokidis, J., Falcone, E. A., Baker, C. S., Burdin, A. M., Clapham, P. J., Ford, J. K. B., Gabriele, C. M., LeDuc, R., Mattila, D. K., Quinn II, T. J., Rojas-Bracho, L., Straley, J. M., Taylor, B. L., Urb?n R. J., Wade, P., Weller, D., Witteveen, B. H. and Yamaguchi, M. (2011). Humpback whale abundance in the North Pacific estimated by photographic capture-recapture with bias correction from simulation studies. Marine Mammal Science, 27: 793-818.