

成績分佈曲線

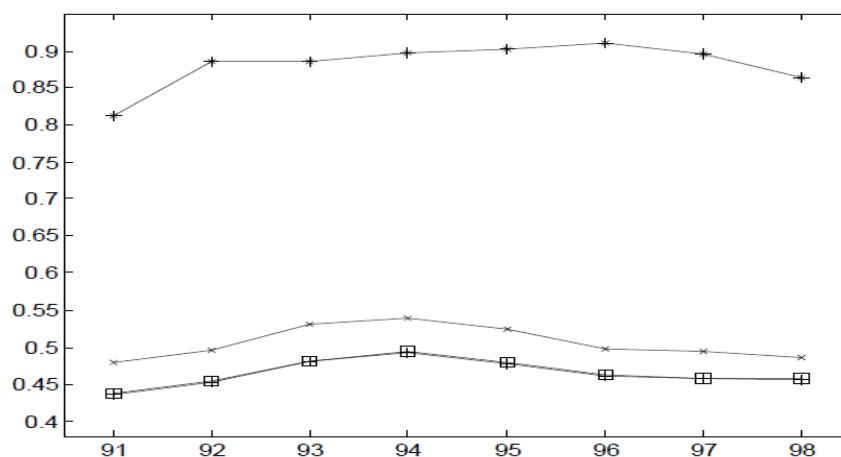
單維彰·98年9月18日

在本刊的今年8月號(第476期),萬芳高中物理科的紀延平老師,發表了一篇〈指考物理真的太簡單、沒有鑑別度嗎?〉的評論。那是回應今年7月2日物理科指考結束後,部分補教界的看法(中國時報將之放在頭版頭條)。針對所評論的問題,紀老師的意見是否定的;我相信他的意見也是大多數高中物理教師的意見。

紀老師的評論,包括三個意見:物理科的成績分佈,不是漂亮的「鐘型曲線」,它不是沒有「鑑別度」,而是僅僅爲了明星大學的選才而服務,而這個現象很可能降低中學生選讀物理的興趣,危及整個國家的自然科學與理工人才之基本培育工作。這些意見引起我個人的進一步興趣,特別是想要明白,爲什麼考試成績的分佈,總被認爲應該是「鐘型曲線」?事實上,近年的數學甲和數學乙的指定考試成績分佈,就經常不是「鐘型」的,而更接近「水平」的;我想要說明,就一份考題而言,水平分佈是比較理想的。以下,我向讀者報告,在這些議題上個人調查和思考的心得。

我使用的所有資料,都是從大考中心的公開網站上取得。其首頁網址是<http://www.ceec.edu.tw>,在[測驗考試]→[指定科目考試]→[統計資料]裡面,可以獲得許多種數據。至於圖,都是我取得數據之後,另外用Matlab軟體畫的。

雖然指考的報考人數連年下降(自95年至98年,報考人數的比例大約是100:92:86:82),但是近年來報考物理、化學、數學甲和數學乙的人數比例,卻大約持平,如圖一。橫軸是年份,縱軸是數學乙(*)、數學甲(x)、化學(□)和物理(+)的考生人數佔當年報考人數的比值。我們看到物理和化學的考生比例幾乎不可區分,而化學科並未傳出長期以來考題太難的批評,或許可以解釋爲:物理科的考生人數比例,並未因爲考題偏難而降低。



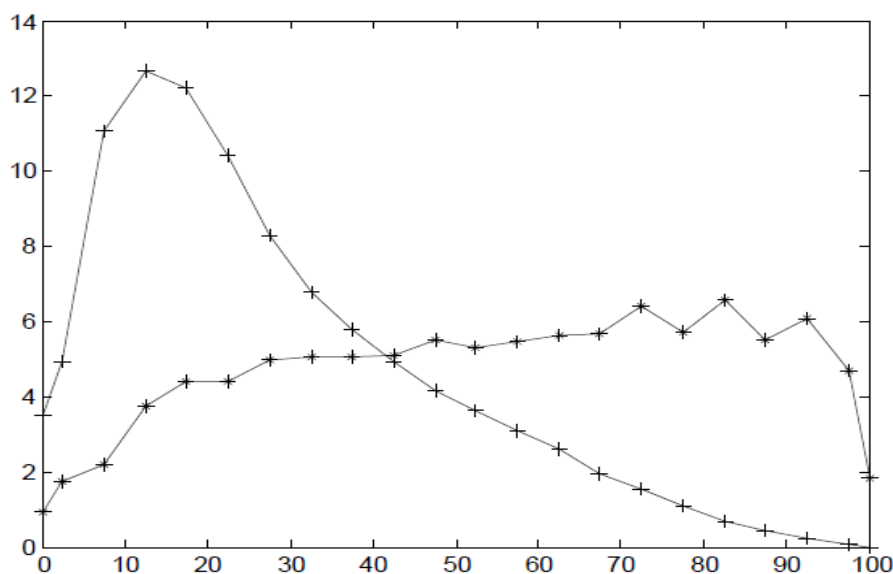
另一方面，我們看到物理和化學的考生比例穩定低於數學甲。如果將物理和化學視為「理工」的同義詞，這似乎意味著數學甲的「市場」比理工稍微大一點。但是，如果將數學甲、物理和化學視為同一類型的「理工」性向（雖然我知道這並不正確），則有一部份具備理工性向的學生（報考數學甲），放棄了物理和化學，這就或許是值得更仔細研究的警訊了。

如果考題變簡單，會不會提高報考人數（也就會提高對應學科的選修人數）？我們沒有統計證據支持或反對這個命題。但是，我相信考生（及招生的大學科系）更在乎的是「鑑別度」，而不是考題的難易。畢竟，全體考生都得到 90 分，與全體考生都得到 10 分，在競爭和分發上，是沒有差別的。而且，就一名教師的直覺而言，考題的難易絕對會影響學習態度；或許這個命題太明顯了，所以乏人做正經八百的調查研究。

所謂有或沒有鑑別度，就是能否用一張考卷，將考生如切香腸般地切成一片一片的，每一片都差不多厚。一張 20 道題目的考卷，最好能像有 20 關的障礙競賽，一關比一關難過。過關的人向前，不過的人留置在兩關之間。到了最後，希望每兩關之間所留置的人數，大約一樣多。

順便說明一下，並非所有考試都以達到「鑑別度」為目標。例如教育研究院籌備處即將在全國抽樣舉行的「學習成就評量」就不是如此，它的目標是探測學習的現況，作為制訂教育政策（諸如課程綱要的修訂、補救教學的經費和重點等）的佐證。像這種考試，就算全體考生都考滿分也不要緊。

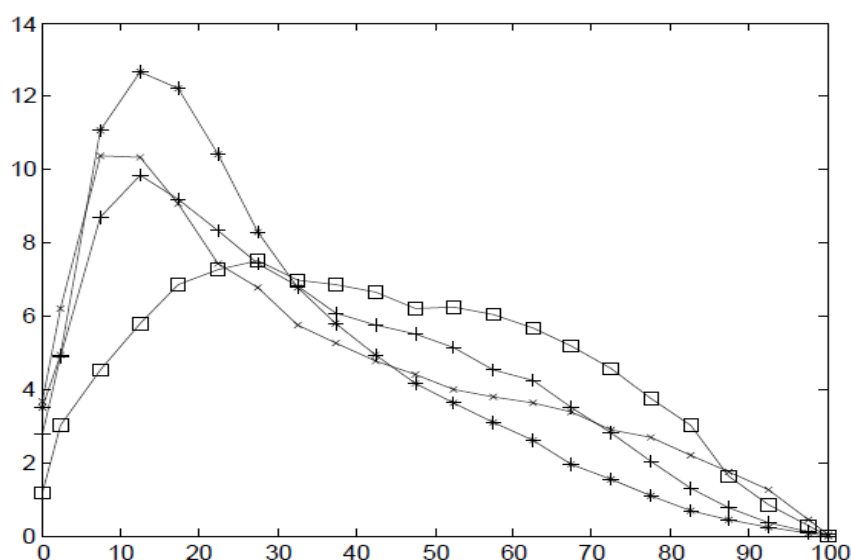
我們用以下方式畫出一份考卷的成績分佈圖：先將分數畫割為 20 段，每段 5 分。例如第一段是 $(0 \leq \text{成績} < 5)$ 的範圍，第二段是 $(5 \leq \text{成績} < 10)$ 的範圍，最後一段是 $(95 \leq \text{成績} \leq 100)$ 的範圍。我們算出每一段成績的考生人數，除以應屆有參加考試的人數（剔除缺考者），乘以 100，得到每段成績的人數百分比。此外，我們又特別畫出得到 0 分或 100 分的人數百分比。圖的橫軸是成績，縱軸是百分比，每條曲線由 22 個點相連而成。圖二裡是 95 年度的物理 (+) 和數學乙 (*) 的成績分佈圖。



由圖二可以觀察到，紀老師所謂物理考題只為明星大學服務的意思。那一年的物理考題，超過 3% 的考生 (大約 1500 人) 抱蛋，而大約 35% 的考生 (大約 18000 人) 擠在 20 分以下的區域，他們的相對能力，並不能被有效地區隔開來。這些學生也交了報名費，他們的潛在學校也是大考中心該要服務的對象 (雖然當初成立的時候，只有 48 所學校出資)。相對地，那一年的數學乙考題，在相當寬的成績範圍內「宛若」一條水平線，也就是每一段 5 分範圍內的考生人數差不多一樣；這些考生的相對能力，就被仔細地區分開了。

我們現在討論的純粹是考卷「鑑別度」的意義與技術，並不討論成績究竟是否等於能力？能力是否可以被鑑別？後者是哲學或道德問題。

觀察 95 (*) 年、96 (x) 年、97 (+) 年和 98 (□) 年度物理考試的成績分佈圖，可以看到曲線漸趨「平坦」，而零分的比例逐年下降，如圖三。可見物理科的命題委員們，很努力地在改進他們的命題技巧。



一幅「平緩」的成績分佈曲線當然就不會是「鐘型」，這似乎違背了許多人對統計的直覺。用一道題目去考學生，如果仔細地評定部分分數 (如數學的計算證明題)，則結果可以像調查身高體重一樣，產生一個鐘型的「常態分佈」。所以，兩道題目就是兩個常態分佈。假設一題難一題簡單，其平均分數分別是 3 分和 8 分。根據統計理論，如果兩個分佈「獨立」，則相加之後仍然是常態分佈，而且平均分數是 $3+8=11$ 分。

如果學生都用猜的，則猜第一題的對錯，與猜第二題的結果，就是「獨立」的。在此情況下，總分的分佈才會接近鐘型。但是，學生做考卷不應該總是用猜的，而是有實力的。如果考題能夠讓學生發揮他的實力，那麼兩題的分數就不該是獨立的：實力強的人，就應該傾向於在每一題拿到高分。所以，經過精心策劃，的確有可能設計一份區分學生實力的考卷。

如此說來，成績分佈接近鐘型的考卷，反而啓人疑竇：我們無法確定，學生全都用猜的呢？還是憑實力寫的。圖四顯示兩條曲線，一條是用電腦模擬一萬名

學生隨機猜答案的成績分佈圖 (*), 假設每人每題的猜對率都是 50%。另一條是 98 年指考的國文科成績分佈圖 (○)。電腦模擬出來的曲線果然是鐘型的。令人驚訝的是, 模擬的曲線與國文科的成績分佈如此接近!

