

統計數字會撒謊(下)

主要資料來源：《赤裸裸的統計學》，Charles Wheelan 著(美)，曹 檳譯，中信出版社，2013。

今天的 1 美元和 50 年前的 1 美元價值是不一樣的，那是因為通貨膨脹。所以如果要以金錢來作比較，不考慮通貨膨脹都是不準確的。經濟學家對此現象就使用了不同的術語，對於沒有就通貨膨脹做出調整的數字就冠以「名義」二字。包括美國勞工統計局在內的許多網站，都提供簡易的通漲計算器，讓人們對不同時期的美元價值進行比較。再以法定的「最低工資」為例，隨著時間的推移，通貨膨脹會逐漸削弱最低工資的購買力。如果商品價格的上漲速度，快於調高最低工資的速度，那麼最低工資的實質價值就會縮水。

統計學的一個重要角色，就是描述數量隨著時間推移所發生的變化，我們繳的稅是不是越來越多？與去年相比，公司的業績有沒有增加？……。我們常用百分率來描述這些變化，因為百分率能給我們相對直觀的感受。百分率雖然不會撒謊，但常會被用來有意的操弄。如果起點非常低，進行百分率的比較就可以呈現「爆炸」式的增長；另一方面，如果一個龐大數額的微小比率，也會是一個很大的數字，反之亦然。「公司全員調薪 3%」，聽上去要比「張三調薪是李四的 20 倍」好受太多了，雖然這二種說法都沒有錯！

只要對一段時間內的數字變化進行比較，就一定離不開一個起點和一個終點，於是就可通過操縱這些點來影響訊息的表達。美國有位教授對共和黨和民主黨操縱數據的伎倆十分清楚，尤其是在軍費開支的問題上，他指出就算是面對完全相同的數據，不同的分析方法也能夠產生不同的效果，既可以用來取悅民主黨的支持者，也不會讓共和黨的擁護者失望。他就準備了二個版本的幻燈片，為共和黨人上課時，就拿出「共和黨版」的，為民主黨人上課時，就拿出「民主黨版」的，裡面的數據完全相同，只是組織數據的方式不同，並配上不同的解讀而已。

人們都希望獎勵並效仿「好」的學校，整頓「差」的學校。可是學校辦學的品質要如何衡量？考核學校和教師最常用的方法就是看學生的統考分數，學生的優異成績是教師和學生最好的金字招牌，糟糕的成績就意味相關的教師應該被辭退，而且越早辭退越好。錯了！如果只看學生考試分數是會鑄成大錯的。不同的學生背景能力是很不一樣的，不管孩子上的是哪所學校，學生父母的教育程度和收入，會對孩子的成績產生不可忽視的影響。那麼，學生學業表現有多少比例要歸功或歸咎於學校？我們還缺乏數據！從小就生活在書香門第、衣食無憂家庭的孩子，一般來說從進幼兒園的第一天起就有可能比別的孩子成績好。有些學生天資平平，雖然教師教得很好，但是學生的成績還是受

限，可是如果沒有這些老師的付出，這些學生的成績就會更糟。因此，學生成績的絕對值對於解答這個問題並無意義，因為我們想要評估的是學生的表現中，有多少是受到了學校和教師影響的這些教學因素。

有人會說這不難啊，只要在入學時對學生安排考試，再將這次考試的成績與入學之後的考試成績進行對比，不是就能判斷學生的學業是進步了還是退步了，並由此來評價學校或教師。可惜，這種方法還是錯誤的。不同能力背景的學生在學習上的進步程度也是不同的。一些學生在領會知識方面就是比其他學生要快，與老師的教學品質是沒有關係的。某一所名校的學生數學成績好，原因可能是老師的教學能力比較強，也可能是該校學生的學習能力強，或是二者兼有。然而學校要提高學生的考試成績，除了改善教學品質還有別的辦法，就是讓那些成績不好的學生「遠離」考場，例如留級或使用其他招數不讓他們拉低全校的總平均即可(操縱統計數字)。

紐約州政府為了讓公眾在選擇醫生時有一個參考，出台了「記分卡」制度，對接受心臟搭橋手術病人的死亡率進行統計。這顯然是一個非常重要的數據，應該是一個好政策，結果卻反而給病人造成了傷害。為什麼？因為降低病人死亡率的最簡單的辦法，就是拒絕為那些病況最嚴重的病人動手術。於是一些本來可以從搭橋手術獲益的病人，最終沒能被安排進行手術，事後的調查證實有 79% 的醫生表示，這個政策或多或少地影響了他們的治療決策。這一看似有用的統計數據存在一個可悲的矛盾，而醫生也只能理性的接受並採取自己的對策，也就是讓那些最需要心臟搭橋的病人遠離手術台。

作為一個統計指標，不僅時常會「攜帶」其他描述性數據可能存在的陷阱，而且就其自身而言，將許多指標數據融合為一個單一的數字，這多少都會對現實和真相產生扭曲。因為任何一個指數對其構成都是十分敏感的，無論是所採用的指標數據變化，還是各個數據權重的變化，都會對指標數據產生影響。如果計算過程的不確定性無法克服，那麼花大力氣將眾多數據壓縮成一個數字只是為了簡便，這樣做是否值得的？

由以上討論，我們知道統計陷阱與數學能力的關係不大。那些看起來令人嘆為觀止的精確計算也會混淆視聽，甚至成為不良動機的掩護。再好的統計學知識也不能遏制不道德的行為，壞人總是清楚知道自己在做什麼，客觀判斷和正直才是關鍵所在！

心得分享：

以國外權威的大學排名為例，它的確能反映高等教育的成果嗎？難道排名在前大學越多的國家就越強，缺乏排名在前大學的國家就很難發展嗎？事實證明也未必！如果一個學校為了排名，花了大量精力與資源去配合評鑑項目來辦學，對學生的學習來說是福還是禍？站在學生的立場，最關注的應該是在學習知識的本身上，大學教育對自己有什麼影響、自己的能力可獲得怎樣的提升，其實自己是最知道的。

為什麼權威的大學排名，評估的指標偏重「輸入」，如錄取率、教職員薪資、教授占比、生師比、每個學生能獲得的資源...等(背後的原因是這些數據比較容易計量)，而在「輸出」方面僅有留級率和畢業率(也是比較容易計量)，完全看不到學生的能力是否提高了、知識是否增長了？正因為排名只看這些輸入的評估指標，就對參與排名的大學產生一個誘因，為了提高自己的評分，把有限的資源投入那些能得到積分的項目，而忽視了大學「培育人才」的基本任務。因此，有些已有知名度的大學現在已經不再「報名參加」排名，不再提供資料給所謂的大學排名機構了。你認為要評估一所大學的辦學品質，可以建議採用什麼作為指標比較適合？

無獨有偶，台灣警政署對於警察各基層單位績效的評估指標之一，就是「破案率」，這看來好像也很有道理，破案率高不是表示警察在認真做事嗎？可是，請大家想想看，治安案件有大有小，要如何「加權」？如果你是基層單位

主管，請問要如何提高自己單位「破案率」的積分？會產生什麼副作用和影響？

同樣的，我們也很熟悉政客或有心人士的操弄：觀光客不來了，不是空氣變好了嗎？新冠死亡率低必然是隱瞞了真實數據，死亡率高正表示未來社會總體免疫力的提高；提高關稅增加了出口國的負擔，多收的關稅可以拿來補貼本國人；炸敵方的醫院，目的是消滅醫院中藏匿的恐怖分子和彈藥……，相同的現象或數據，以不的角度解釋可以完全不同，本文就是最好的提示。