

## 統計數字會撒謊(上)

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。

雖然統計學是紮根數學，而數學又是一門準確性的學科，但使用統計學來描述複雜現象的過程，並不是精確無誤的，這就為掩蓋真相創造了大量空間。描述同一事物本來就有許多種不同的方式，那怎麼選擇就會影響別人對此事的印象，更別說一些有心人士會用光鮮亮麗的事實和數據，支持真假存疑或完全不能成立的結論。

「精確」和「準確」是同一回事嗎？不是！這二個詞是不可以相互替代的。「精確」是指我們描述事物的精度，如距離目標 41.6 公里，就要比大約 40 公里精確，要比「相當長的一段路」更為精確。但是如果沒有指明方向，那還是「不準確」的，把精確和準確混為一談，是要付出代價的。如果一個答案是準確的，在這個基礎上當然是越精確越好，但如果答案一開始就不是準確的，那麼再精確也毫無意義。

在 2008 年金融危機爆發前，華爾街的許多風險管理模型都非常精確。「風險值」這個概念，讓這些公司將不同情況下的可能損失精確量化，然而這些模型中有關全球市場可能會發生的風險假設確是錯的，因而精確計算出來的結論根本不準確，結果不僅苦了華爾街，更是把全球經濟都拖下了水。

再舉一個例子，下面這兩個陳述句都是有憑有據的，數據也都沒有造假。你能夠看出差別嗎？

- 我們的教育水準正在變得越來越糟！2013 年有 6 成學校的考試成績低於 2012 年。
- 我們的教育水準正在變得越來越好！2013 年有 8 成學生的考試成績高於 2012 年。

(提示：不是所有學校的學生人數都是一樣的。)

請仔細看一下這貌似相互矛盾的陳述，你會發現原來一個是說「學校」，另一說的是「學生」。如果成績上升的學生是來自規模非常大的學校，那麼這二個矛盾的說法是有可能同時發生的。

請大家再比較一下，世界各地人民的收入不均衡現象，因為全球化的到來

是改善了，還是惡化了？一種理解是，全球化加速了現有的收入不均狀況。1980 年時，以人均 GDP 為參考值的富裕國家，在之後的 20 年間的增長速度超過了貧困國家。正是因為富者更富，說明了貿易、外包、外國投資以及其他全球化的組成部分，淪為發達國家擴大經濟霸權的工具。可是，同樣的數據，也可以另一種完全不同的方式來解讀，那就是我們不要針對「窮國」而是針對「窮人」。世界上有絕大多數的人生活在中國和印度這二個超過 10 多億人口的大國，在 1980 年，這二個大國都處於相對貧窮的發展階段。但是，在過去的幾十年裡，中國和印度的經濟都經歷了高速發展，這很大程度上要歸功於，它們與世界上其他國家日益加深的經濟一體化。如果目的是改善人類的窮困，在衡量全球化給全世界窮人帶來的影響時，將中國 13 億人口和毛里求斯的 130 萬人口當成比重相同的國家看待，當然是不合理的。因此，上述例子的分析對象應該是人，而不是國家，「全球化」的確有利於全世界的窮人，全球不平等的現象正在迅速減少。

美國二家電信業巨頭美國電話電報公司和威瑞森電信陷入了一場廣告之爭。從邏輯上來說，要比較這二家公司孰好孰壞，只要看它們各自網路的覆蓋率和通訊品質就行了。二家公司在衡量此一看不見、摸不著的需求時，採取了不同的分析指標。威瑞森公司發動了一場聲勢浩大的廣告戰，大力宣傳其無所不在的網路覆蓋範圍，對比美國電話電報公司相對零碎的地理覆蓋。美國電話電報公司則發動了廣告反擊戰，宣稱該公司能夠滿足 97% 美國人的通信需求。注意這裡的用詞是「美國人」而非「美國」。二家公司提出的數據都是真實的狀況並未造假，只是重點不同。

美國前總統小布希在任時推動一項減稅政策，主管部門說會有 9,200 萬美國人受惠，人均減稅額超過 1,000 美元。《紐約時報》評論說：「數據本身並沒有撒謊，只不過有些數據沒有發出聲音罷了。」是的，的確有 9,200 萬人受惠，但是在減稅政策下雖然平均少繳了 1,083 美元，但中位數還不到 100 美元。這表示只有數量相對少的巨富們才有資格享受大額減稅，正是這些人拉高了平均值。

然而，中位數也同樣會掩蓋事實真相。知名的進化生物學家古爾德曾被診斷出患了某種癌症，她的半數預期壽命(中位數)只有 8 個月，但 20 年過去了，古爾德才死於另一種不相關的癌症。古爾德還因而寫過一篇〈中位數不等於真訊息〉的文章。中位數是告訴我們有 1/2 的病人活不到 8 個月，但另外的 1/2 則有可能可以活到老。