

心得

天下文化
New Culture

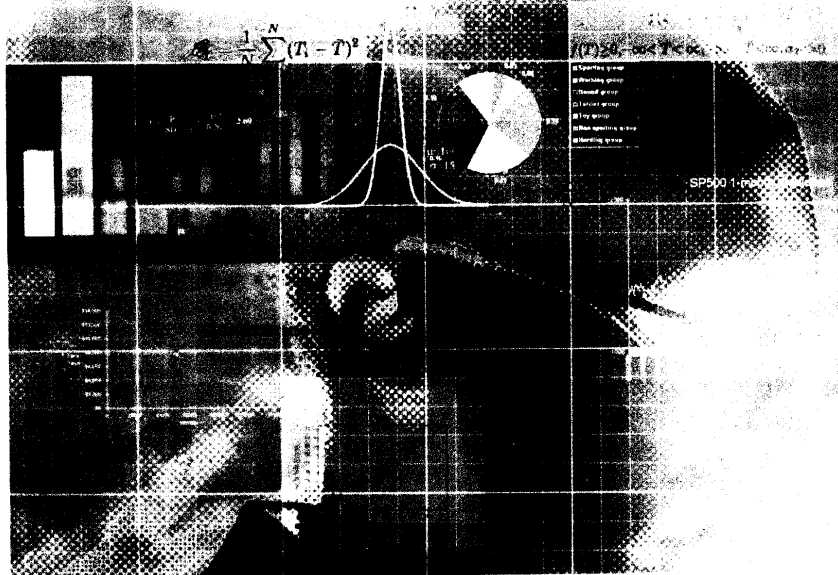
統計

by David Salsburg 葉偉文 譯

改變了世界

The Lady Tasting Tea

How Statistics Revolutionized Science in the Twentieth Century



施尚文



看完此書，我有一個很大的疑惑，爲什麼一本關於統計學的書沒有出現任何圖表與函數和數字，在序文的部分作者有說：此書沒有很大的野心利用數學公式與符號來說明統計世界的數學模型。書皮部分，范斯丁醫學博士也提到：全書刻意避開複雜的數學符號及理論，讀起來一氣呵成。

但就是因爲沒有數學符號與圖表，讀的非常辛苦，記得國三下學期數學——資料整理與機率，有直／橫條圖，圓餅圖，長條圖，老師說這是統計最基礎的部分，就因爲這些圖表使得統計平易近人。此書章節提到“斜率分布”“T檢定”“鍾形曲線”“平賭序列”……，因爲不使用圖表，作者使用大量的文字敘述說明解釋，讀起來反而略爲艱澀，難以理解，實在是一大缺陷。

但不能就如此否定整本書，作者的文筆與翻譯者的功力，許多的概念在一行行的中文字中，解釋的清清楚楚，否則學校老師也不會把它列入書單內。此外，我對於心得的方向以一個章節一個章節逐次寫出，選擇特定感觸較深部分。在此坦承，並沒有將全書完全看懂，畢竟這屬於高中及大學課程的部分。只能盡個人最大的能力來理解。

在第二章 偏斜分布〈相關與回歸〉提到高騰的理論：向平均數回歸：非常高的父母所生的孩子，往往比父母矮些；而非常矮的雙親所生的孩子，則往往比父母高。我對於這個理論很有興趣，看過篇醫學報導（聯合晚報內）指出，依照現在台灣小孩營養充足的條件下，男孩的身高大約介於父母身高平均再加10。我媽常說如果她再高一些的話，相信我會比現在還高。如果身高很高的父母，他們孩子還會再往上長嗎？那篇醫學報導有考慮到嗎？還是把特例刪除？或是台灣沒有很高的父



母？越往深入處想，越覺得有許多疑點未顯現。大概只有看完原始研究報告才能理解。

同章，皮爾生有個想法：我們可以把實驗結果看成許多分散的數字，更精確的說法就是當成一組數字分布（distribution of numbers）。這讓我想起國中的理化實驗，當時想再加入另一個變數，但產生出的圖表將是一個三度空間，一個個數值將會呈現如同正方體對角線的斜直線，但已超出課程範圍，又加上自己能力不夠而作罷，希望高中有機會再接觸。

綜觀整本書，發現科學界有一個奇怪的現象，一個偉大的科學家發現（或發明）一或數個偉大的定理或公式，往往不能接受另一位科學家推翻他的理論，就像在第五章費雪所說：若有人告訴他們，十年前所學的東西應該稍微修正，他們大多能接受。不過也有少數的人不能接受這一點，就好像隱私被揭露，或是自己的私密空間被人惡意闖入；他們的反應會很激烈，而且毫不留情。就像皮爾生攻擊費雪，費雪攻擊尼曼，似乎費雪也是他自己口中的“少數人。”我可以了解道為什麼當初牛頓的光學論文被虎克批評後不再發表論文；高斯不公佈他的研究：三角形內角和小於 180° ，擔心公佈後引起學界永無止境班的辯論。甚至有一個可怕傳說：當希伯塞斯（Hippasus）發現正方形對角線不可公度（也就是無理數），建立在有理數上的畢達哥拉斯學派學者感到非常生氣，而把他帶到遙遠的地中海後將他扔入海中淹死。這種科學界常發生的奇怪現象，我十分不解。

在第八章提到“半致死劑量”又說離50%越遠，就得進行更大規模的實驗，例如25%，80%，我有很大的疑問，為什麼離一半值愈遠，實驗規模愈大，

對我而言， $1/2$ 是一個很模糊的地帶，換言之，是很輕鬆的數值，同樣的數值 $1/4$ 或 $3/4$ 也是很輕鬆的數值，為什麼 $1/2$ 就是一個較準確的估計值，有機會想看看計算半致死劑量的電腦程式。

第十章提到“混沌理論”，一個對初始條件非常敏感的數學函數；經過多次的迭代之後，會造成迥然不同的結果。我完全不能接受這論調，尤其是那演講題目：巴西一隻蝴蝶翅膀的拍動，會是美國德州龍捲風的成因嗎？當然不可能，乾脆我現在用力跳一下，來討論台灣明天會不會有大地震，一樣地無聊。依我學生的觀點這是明顯的鑽牛角尖的情況，在統計時因為特殊值的影響而無止境地研究下去，到最後產生一個偏曲的理論。

在注釋的地方看到凱因斯本身為哲學博士，且提出的論文為〈機率的論述〉令我非常驚訝，在國中唯一提到他的一次於歷史課中，經濟大蕭條時他提出政府應適時地插手控制經濟，使其維持一平衡標準。不過仔細想想並無不有道理，本身研究哲學就有機會接觸機率（我覺得稱的上是另一形態的哲學），再因為當時的世界背景，使用統計來探索經濟動態，畢竟機率與統計可用於許多層面，如：生物、生態環境、社會、流行、政治……等，當然還有數學，這就是個將一學術延伸推至另一方向的例子。

貝氏異論的概念非常有趣，先看結果，在反推回去。此書中康菲德爾的「追溯性研究」作者雖然沒有與之搭上線，但“由疾病並開始，追溯回去看看哪些條件”的概念無異是其概念，在解題時，也可以先看解答，再推敲最快速的解法；我想這也是歷史學家最常使用的方法之一，已獲得一個不知名的古物，再利用上面的圖騰

或整體的造型來判斷屬於哪一朝代，甚至可以利用上面的文字來補齊歷史的斷層，其實在現在的社會上，這種逆推的方法已普遍使用，在當時無法被統計學術界接受實在奇怪，或許是統計的概念屬於另一種邏輯，更加吸引我接觸它。有一部分人說統計學是數學中最複雜的部分，我不這麼認為，但統計一定是需要某種特殊的思考來研究的學科。

看完整本書，有種莫名的感嘆。在科莫格洛夫 (Andrei N. Kolmogorov) 的部分，我不禁要嘆口氣，他在五歲時發現最小的 k 個^奇基數的和，正好等於 k 的平方。我已經一五歲了，才第一次接觸這數值關係，而且是在課外讀物上，為什麼台灣國中、小學教科書在^奇基偶數部分沒有提到？甚至有點無奈：為什麼我不夠聰明而沒發現……而東歐或蘇聯的統計學家在 1930 年代因二次世界大戰與史達林的操控，錯失與其他國家的學者們研討的機會，致而使的許多優秀的研究結果延後發表甚至消失，或是說與所謂的“黨中央計畫與馬列理論”相抵觸，導致數理統計的應用受到壓抑。

第十八章德國醫師科霍提出一組必要公設來檢定特定的疾病是否由某種傳染原引起，滿有趣的：1. 凡是當傳染原可以被培養出來，疾病必會發生 2. 只要沒有疾病存在，一定是傳染原沒有培養出來 3. 若傳染原被消滅，疾病就會消失。若以非傳染性疾病也說的通：遺傳性疾病中，若隱性的異常遺傳基因顯現，症狀必會出現，符合 1.2.3.；可是因環境和不良的生活方式所引起的疾病就不一定了：若一個人常嚼檳榔，很可能罹患口腔癌，但不是絕對的；某個人知道自己因為嚼檳榔罹患口腔癌而停止嚼檳榔，口腔癌並不會因此消失，不符合 1.3. 點。傳染性疾病而言，如果一個人抵抗力強，且遇到的非 AIDS 等世紀絕症，通常 1. 不成立。所以我個人判斷，2. 在於任

一種疾病都說的通。1.3.需視疾病而定。

第二十二章杜奇說：用四條協線和一條交叉線代表五個是個笨法子。沒錯，我在畫時也常常數錯，他提議以畫正方形的形式表示“十”，我覺得也是個笨法子，在畫3時只點了三個點如果不是很清晰時，會誤認為是汙點，再快速計算時更是如此。倒不如使用“正”字記號是最佳的，可快速登記又不會有爭議。其實老外應該學學些中國字可增加生活的便利性。

許多機率與統計的專有名詞融入現代人平時生活中，就因為生活化，導致非學過統計學的人誤用。例如我跟朋友約定會面，他回答我他有75%的機會會到。我差點沒昏倒，什麼叫「75%會到」，會來就會來，不會來就不會來，答案只有為真或不為真，並不屬於機率的範疇。我對機率使用在氣象報導最為詬病，當氣象播報員說：「明日台北降雨機率為95%」他的意涵是：明日台北24小時內有95%的時間在下雨，還是：明日台北有95%的地區會淋到雨，或是：明日台北出門的人中有95%要撐著傘，或是就表面的意思：台北有95%的機會下雨，可是會下雨就會下雨，不會下雨就不會下雨，是不容許灰色地帶的，那真正的意思到底是什麼？

看完本書，雖然對其中的許多例子並不認同，但仍學到了很多數學與非數學的事物與定理，也了解到統計史的發展，身為二十一世紀的高中生，期許自己因讀過此書而在未來的高中三年對於統計有更好的吸收力與融會貫通，不負作者寫了將近198000個字的苦心。