

統計思維

黃文璋

國立高雄大學應用數學系

前言

美國著名的小說家馬克吐溫(Mark Twain, 1835-1910), 在1907年的自傳裡, 引用曾任英國首相的迪斯雷利(Benjamin Disraeli, 1804-1881)的話:

There are three kinds of lies: lies, damned lies, and statistics (有三種謊言: 謊言, 可惡的謊言, 及統計)。

由於馬克吐溫的高知名度, 這句話因他說了之後, 便廣為流傳了。

大家都學過多年數學, 對於為什麼要學數學可能會較清楚。原因之一當然是生活上, 及專業上, 會用到一些數學, 也就是數學可視為一種工具。而對一個數學學通的人, 會有什麼特質也還算清楚。不外是較有邏輯, 以及計算較精準等, 大抵是人們所喜歡的特質。

近年來, 中學數學教科書中, 統計的份量增加不少。進入大學後, 不少學系也都要學統計。看起來統計學似乎愈來愈重要。但我們一方面看到有人做決策時, 非有統計不可, 把統計當護身符。卻也會看到有如馬克吐溫者, 對統計嗤之以鼻。即使在學術界, 有些人以為統計不過就是數學; 但有些統計學者, 會一再強調統計與數學是完全不一樣的。看起來統計的內涵似乎不易令人掌握。比如, 我們可能知道什麼是很有經濟頭腦, 什麼

是很有文學細胞，以及什麼是很有音樂素養。那什麼是很有統計頭腦？統計細胞？或統計素養？就不易講得明白了。

底下我們試圖藉由闡釋統計學裡的思維方式，來略化解前述關於統計內涵的疑惑。首先來看，統計學究竟在做什麼？

如同其他科學的學門，統計學也是要告訴人們一些結果。對待釐清的事，給出一些推論。我們可以簡單地說，統計學裡所能達到的是：

1. 允許誤差下的機率保證，
2. 允許誤差下的無罪推定。

在數學裡所探討的多半是**必然性**的問題。當它說1就是1，不會有些微誤差。而一命題一旦被證明是對的，問題就底定，不會有例外，除非你能找出證明的漏洞。而統計裡，是在處理**隨機性**的問題。它允許**誤差**，沒有誤差反令人懷疑其中有假。統計裡也會拍胸脯保證，但它的保證，都是機率式的。而且通常所能保證的機率，不但不是百分之百，還附有誤差。金庸在“倚天屠龍記”一書裡，塑造出一個布袋和尚“說不得”。統計裡則處處是“說不準”。例如，宣稱有百分之九十五的機率，某飲料的容量，介於326cc至331cc間，就是一典型的統計保證。雖有人說統計與算命的工作性質類似。但在統計裡，少有鐵口直斷的，總顯出有點保留的樣子。

我們很少企圖經由統計，去證明那一件事一定是對的。你想探索真相嗎？真相就留給上帝吧！要知在隨機世界中，真相常是難以大白。一切都是假設，只看你接受那一個而已。而接受，就如在教堂裡，當新娘點頭說“我願意”，並不表示這位新郎就真正是最適合她的。只不過是“目前她願意接受”。同樣地，在統計裡接受不表示為真，拒絕也不表示為偽。而接受或拒絕，採用的是類似刑事訴訟法（第154條）裡，**無罪推定**的精神。只是不像法庭上，法官敲擊法槌後所做的宣示之不容置疑。統計學家的判定，往往還會給出誤差。

統計學裡的某一方法，常對應人們的某種思維方式。由於人們有不同的思維方式，皆有其道理，各有適用的時機，因此也就有種種的統計方法。這些方法的優劣，有時是可以比較的，但仍是允許誤差下的機率式比

較。

機率及誤差，構成統計裡的思維之兩大支柱。因而發展出統計學裡所著重的幾項要點，即善用資訊，了解變異，相信機率，合理估計，無罪推定，及紙上談兵等。我們將分別來說明。

1 善用資訊

在柯南道爾(Conan Doyle)著的“桐山毛櫟山莊”(The Adventure of the Copper Beeches)一書裡，福爾摩斯(Sherlock Holmes)說：

“Data! Data! Data!” he cried impatiently. “I can’t make bricks without clay.”

沒有規矩不能成方圓，沒有黏土不能做磚，沒有資料(data)則無法做決策。福爾摩斯可以依命案現場的一些蛛絲馬跡，推測凶嫌可能慣用左手，或可能經過一片果園。算命看相者，所仰賴的也是資料。收集很多不同的面相及八字等的命運，當“閱人多矣”後，自然容易依據人的面相等，分析其前程。那些善於看透人性者，不也是閱人多矣嗎？做決策要有資料，每一項資料，都可能是有用的資訊。統計學家的本事要能發揮，就得善用資訊。因此對於統計學家，資料有如老鼠所愛之大米。

有人敲門，是男是女？大約各半吧！這是因基於社會上男女比例差不多是1比1。但如果你從門縫看到來者穿裙子，就會覺得極可能是女生。因在經驗裡，少有男生穿裙子的。如果你知道來者是個數學系的學生，那仍依你的經驗，敲門者是男生的機率可能約為8成。原來男女各2分之1的機率，是在沒有其它資訊下的假設。一旦有更多的資訊，此假設便不必然仍存在。如果這些你可以接受，那也應能接受在統計裡，機率值是會變的，這就是所謂條件機率。

只是有些明明是有用的資訊，卻不見得容易看出。在某公園中，你新認識一個朋友。聊啊聊，他指著遠方說“我太太與小孩都在那兒。”你一看，有個媽媽跟兩個小孩在一起，其中一個是女孩，另一個小孩蹲在地下抱著狗，無法得知是男是女。如果仍假設生男生女的機會各半，那蹲著的

小孩，是男是女的機率，不應也就各為 $\frac{1}{2}$ 嗎？你大概不會覺得這是什麼難題吧！只是這裡面卻有些玄機。一家庭兩個小孩的性別有男男，男女，女男及女女等4種可能（其中男女表老大為男，老二為女，餘類推），機率各為 $\frac{1}{4}$ 。今已知有1女孩，便知必為男男之外的3種可能，機率各為 $\frac{1}{3}$ ，而其中有兩種情況此家庭有男孩。故抱著狗的那小孩是男孩之機率為 $\frac{2}{3}$ ，是女孩之機率為 $\frac{1}{3}$ 。如果抱著狗的那個小孩明顯地比較小（是老二），則兩個小孩的性別成為女男及女女兩種可能。故抱著狗的那個小孩，是男是女的機率各為 $\frac{1}{2}$ 。你可能沒想到吧！知道一小孩之排序（老大或老二），居然會影響對其性別之判斷。看來不要輕易放過任何資訊。

在上例中，如果有進一步的資訊，比方說較小的孩子一向比較愛抱狗，或男孩一向比女孩愛抱狗，那對於抱狗小孩為男或女之判斷，也將隨之而變。而如果由之前的聊天中，你獲知朋友兩小孩的性別一樣，則抱狗的小孩不必猜了，就是女孩。

機率值會變，是機率的一特性。視新的資訊產生，對一事件機率之判斷，也宜隨之而變。這原本是合理的，隨機應變是也。如果我們說人不宜墨守成法，不能不知合變，那就也該了解能善用資訊，以隨機應變的重要。我們的決策，應有如唐朝柳宗元在“袁家渴記”中所寫的“搖颺葳蕤，與時推移”。

稀鬆平常的草，進了牛的肚子後，卻能產出頗富營養的牛奶。巧妙各有不同，資料進了統計學家手中，有各種科學的方法，將其轉化為有用的資訊，繼而做出決策。我們要學習這些統計方法，以協助我們做出較好的決策。但更要緊的是，我們要讓自己的思考是具隨機性的，有福爾摩斯的敏銳，留意任何不起眼的資料。因表面上看起來似乎沒什麼用的資料，有時對決策的影響，能產生如“蝴蝶效應”(butterfly effect)般的具大。

我們固然常想多得到一些資訊，但有時又想儘量減少資訊的透露出去。例如，號碼鎖、提款卡，或上網的密碼等，都會希望所設定的密碼，要愈難被猜中。如何減少資訊？不少人在製密碼時，傾向挑選與自己相關的數字或文字。如生日，女兒名字等，以較易記憶。但這些資料，他人其

實極易獲得，往往試了幾次便猜中了。又如，有人以1, 2, 3, 4, 5, 6為提款卡密碼，以為萬無一失，別人豈那麼巧也想到？偏偏人同此心，拾獲(或偷取)其提款卡者，有時第一個試的就是這組號碼。香港曾有實例發生。另外，你看過艾爾帕西諾(Al Pacino)主演的“針鋒相對”(Insomnia, 2002)嗎？他將一把槍藏在空調排氣口裡，以為應很隱密，結果當然是被找到。沒有破不了的密碼，端看要花多少時間，我們只能儘量降低密碼被破解之機率。運氣不論，有什麼好方法製密碼？

一袋中有9個白球及1個黑球，隨機取1個球，你猜取中什麼？當然要說白球，猜中的機率是0.9。如果袋中有5個白球及5個黑球呢？就比較難猜了，猜中的機率是0.5。樂透彩開獎，一般的設計是頭獎號碼隨機產生，也就是讓那幾百萬，或上千萬種組合的號碼，每一組合出現的機率都一樣，且每次開獎與以前都毫無關連。則即使收集很多以往開出的號碼，對下一期所開出號碼的預測，並無任何幫助。換句話說，沒有明牌是最難猜的。

所以，你要製密碼，就宜用抽籤或亂數表產生。要知一般而言，人的天性是沒有隨機性的。老師上課點學生號碼上台，他以為是隨機地點，其實不自覺中，就是有些偏愛的號碼。一學期下來，往往輕易被學生識破。以抽籤是較好的方式，學生較難猜誰會被點中。同理要藏東西，不妨將能藏東西的地點編號，隨機抽一號。依這樣的方式，所提供的資訊最少。

2 了解變異

男生1人重32公斤，10人共重幾公斤？這是小學裡的數學題目。長大些後，你知道即使是同年級的男生，也不會每個人都同重。又不只是不同的人，同一個人投擲標槍，同一個人跑步，同一個人的體重，甚至應為一常數的二定點之間距，每次量測可能都有些差異。變異！你開始注意到變異，各種數量不再全然是常數。逐漸地，你也開始有了隨機的概念。你知道考試總有些不可測的因素(就姑且說是運氣吧)。你也知道有些人命不錯，含著金湯匙出世。但你又聽說命好不如運好，運自然是屬於隨機的。

宇宙的運轉，有**必然性**與**隨機性**交錯著進行。例如，我們知道哈雷慧

星每76年接近地球一次(這是必然性)。只是雖能知道76年後的事,但明天會不會下雨?就不是那麼確定了(隨機性)。又如,將手上的銅板鬆開,在中學物理課程裡學過,如果忽略空氣阻力,則在高度固定下,銅板落地所需時間,是個定值。但落地後那一面朝上?就無法預知了。

根據我們的經驗,也有醫學上的依據,龍生龍鳳生鳳,較高的人生出的小孩往往也較高,但仍會有些變異。龍兄鼠弟,例外一向不少。所謂一樹之果有酸甜之別,一母之子有賢愚之分。努力就有好收穫?聰明的人成績就較好?諸如努力,才智及遺傳等,對一個人的未來,通常很有決定性的影響。但世事多變,也不盡然就都如此。意外、豈有此理之事屢有發生。造物者讓人們對未來,知道大致會發生那些事,及如何發生。但又不想讓人們對於未來,能完全掌握。要知在隨機世界裡,必然性使人們願意事先好好準備,而隨機性則使人們對未來,充滿著盼望與戒慎恐懼。光有必然性的世界,毫無變異,對未來缺乏盼望,將讓人們少了努力的動機。而光有隨機性的世界,只靠運氣,將令人失去積極認真的企圖心。三分天注定,五分靠打拼,兩分靠運氣。我們必須敬佩造物者這樣的設計。

由於變異無可避免的存在,我們所能做的,便是要了解變異,有時還要設法減少變異。以抽樣調查某產品製造的良品率為例。一個常見的作法是,抽出的產品若為良品,以1表示,不良品則以0表示,如此得到一個像是1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, ...的數列。再以抽至第 n 次後,總共得到1的個數,除以 n ,當做良品率之近似值。也就是以平均1出現的次數,來估計良品率。人們習於這樣做,而且似乎也知道,如果想要估計夠精準,也就是想讓樣本的平均良品率,與實際的良品率之差異小些,則樣本數 n 便要較大。付出較多的代價(取樣大),以換得更大的準確性(變異小些),這是合理的。諸如銅板出現正面的機率,湖裡某種魚所佔之比率,選民對某候選人的支持度等,這類0、1;正、反;是、否;及成功、失敗等,兩個結果的現象,想知道其中一種現象出現的機率到底為何?都可用上述“平均成功率”來表示。反之,如果你原本便知良品率,現假若有人並未實際抽樣,則檢視那一串0, 1的數列,並不難讓你懷疑很可能是捏造出來的。造假者不易以“隨機”一詞來強辯。這是我們的幸運,雖世事多變,但萬物有

常。在隨機世界中，是存在所謂**隨機法則**的。法則之一，便是前述平均法的理論依據——**大數法則**(law of large numbers)。原來那個有如無理數不循環的小數部分，看似沒有規律的0, 1數列，其實被大數法則規範0, 1出現的情況。但既然是隨機數列，則不論 n 多大，都不能保證前述平均值，就剛好等於良品率。那誤差究竟有多大？

在數學中，常在求近似值。當以一多項式，來近似一比較複雜的函數時，必須要能給出誤差大小，否則這種近似的用途便不大。射飛標，有時偏右有時偏左，不能只說平均會命中紅心。一支職棒球隊的平均年薪雖很高，但除了少數幾位身價超高者外，很多球員可能只得溫飽。還要知道薪水的變異，才能對整支球隊的薪資結構，有較清晰的概念。另一個重要的隨機法則——**中央極限定理**(central limit theorem)，便告訴我們，在一些並不太嚴格的條件下，量測所得之誤差有**常態分佈**(normal distribution)。與數學中的誤差不同，此處誤差大小是隨機的。但誤差之散佈情形，則能描述。

上述二隨機法則要能適用，所抽取的樣本，得是較一般的，這樣才能有代表性，不至於太偏差。在統計裡，常在做預測、做估計。本質上是在做**以偏概全**的事。雖偏卻能概全，這是統計學家的本領。但如果樣本實在太偏差，那就真是以管窺天，見不到全貌了。更明確地說，各次取樣要彼此沒有關連，即這些樣本須**相互獨立**。而且這些不同的樣本要**分佈相同**。以估計銅板出現正面的機率為例，不能每次投擲，只是往前輕輕一丟，這樣每次大約都得到相同的面(不獨立。另外，也不能每次所用的銅板，出現正面的機率不同(分佈不同)。

但**獨立且分佈相同**的假設容不容易辦到呢？如果是投擲銅板，此假設大致會成立。如果要藉由隨機地投擲芝麻，落進一圓中的芝麻數，與落進其外切圓的芝麻數之比值，來估計圓周率 π ，可能也大致還可以。但對於要處理人的問題，就不是很容易辦到了。一方面不同的人之間有差異，一方面人可不像銅板般，會馴服地讓你重覆實驗。以民意調查為例，人不見得會誠實回答問題，甚至還會改變主意。所以實際應用統計方法時，要更謹慎，否則造成的變異，會比預期的大很多。

事實上，即使在弱一些的條件下，大數法則及中央極限定理也仍適用。在大量的觀測後，大數法則指出，所得到的平均值，會接近該得到的值。平均值會在此該得到的值附近波動。至於誤差，為一隨機的量。中央極限定理給出，誤差差不多有**常態分佈**。科學上常要做量測。有史以來三大數學家之一的高斯(Carl Friedrich Gauss, 1777-1855)，曾研究**誤差理論**。在一些假設下，他亦導出量測的誤差有常態分佈。因此常態分佈又稱**高斯分佈**(Gaussian distribution)。德國現在用歐元，往昔用馬克。德國10馬克，是以高斯為人像。高斯在數學上有諸多重要成就，但在10馬克上陪伴高斯的，不是其他，就是一常態分佈的曲線。可見此分佈不只在統計，甚至在科學上之重要。

最後，必須要了解的是，那一串0, 1數列，1出現的“相對頻率”(即出現次數除以觀測次數 n)，會接近1出現的機率。但此不表1出現的頻率(即出現總次數)，會接近我們所期望1該出現的次數。以投擲一公正銅板為例。對於 n 為2，與 n 為10，你認為是前者還是後者較易出現正、反面數相同？答案是前者，發生之機率為2分之1，後者發生之機率為1,024分之252，約為0.246，不到4分之1。若 n 為10,000，此機率將更小，約為0.008。即若 n 愈大，出現之正面數，將在投擲數之半附近一更大的範圍內波動(變異增加)，愈易偏離所期望的投擲數之半。大數法則是針對平均，而非和。也就是所得之正面數除以投擲數，才會愈來愈接近(變異減小) $1/2$ 。很多人不明白這點，看到樂透彩已經開了這麼多期，怎麼各號碼出現的頻率差異愈來愈大，以為真有氣比較旺的號碼，或真有冷門號碼。大數法則是告訴我們，當開的期數愈多(n 愈大)，任一號碼出現的**相對頻率**，會愈來愈接近該號碼出現的機率值。相對頻率的變異會變小，至於頻率的變異則會變大。而且這還不是專業的說法。比較正確的講法是，當開的期數愈多時，任一號碼出現的相對頻率，與該號碼出現的機率值之差異，會很小的機率將很大(即接近1)。這是隨著 n 的增大，變異會愈來愈小的意思。

3 相信機率

有法國牛頓之稱的數學家拉普拉斯(Pierre-Simon Laplace, 1749-1827)曾說“大部分生活中最重要的疑問,都只是機率的問題”。在隨機世界裡,機率一詞大家琅琅上口,但一般人是否真了解機率的涵意呢?

曾在網路上看到回答生雙胞胎的機率之底下一段話,我們一字不改:

機率是統計上騙人的東西,許多事情要重複做100次才有機率可言。懷孕不可能100次,每次懷孕生雙胞胎機率是 $1/89$,但單次懷孕生雙胞胎機率若不是0%,就是100%。就好像問我,50元銅幣丟到地上一個,是蘭花機率有多少?事實上,50元銅幣丟到地上,不是總統府,就是蘭花。如果丟到地上100次,那麼機率就會接近50%。如果丟到地上1次,蘭花的機率,若不是0%,就是100%。

回答者對機率的解釋,雖不正確,但說不定是不少人對機率的認知。

機率的意義是什麼?在諸如投擲骰子,或抽籤時,我們常以“相同的可能性”來解釋機率。即骰子的6個面,每個面出現機率皆認為是 $\frac{1}{6}$;如果有 n 隻籤,每隻被抽中的機率,皆視為 $\frac{1}{n}$ 。此一解釋,在日常生活中,或賭場裡,還蠻適用的。當沒有其他資訊時,常假設每一可能的結果,發生之機率都一樣。第二種方式,是以相對頻率來解釋機率。如果一位職棒選手,過去的打擊率是0.327,你知道打擊率就是安打數除以打擊次數,則一旦他站上打擊位置,你所認為他會擊出安打的機率,大抵就是0.327。這一種常見的對機率之解釋法,還蠻客觀的。其背後之理論基礎,就是大數法則。針對的現象,是可以重覆觀測的。此外,對於世界盃足球賽,巴西封王機率為1成,追上某一女孩之機率為6成等,在這類情況,並無法重覆觀測。少有女孩子會讓你追不上後,換個方法重追一次,再重追,……。因此其中的機率,是主觀的。當然有時主觀機率的產生,也可能是依據過去一些客觀的資料。

上述三種對機率的解釋法,有時會交錯使用,或彼此相驗證。例如,雖原先主觀地認為骰子為公正,但投擲夠多次後,依各面出現的相對頻率,

說不定會調整對各面機率之看法。

對機率的解釋，並不只前述那三種。在數學上，對於骰子，我們也可就指定每個面會出現的機率值。這些機率值當然不能是負的，而且加起來要是1。這是所謂以**公理化的方式**引進機率，在此不多討論。無論如何，機率絕非“騙人的東西”。要知事件在發生前才有機率可言。巴西隊一旦奪冠，再去說該隊封王之機率為何，已無意義。奪冠機率再低，贏了就是老天眷顧。奪冠機率再高，輸了只能徒呼負負。而觀測一事件，結果是“不發生”，或“發生”，而非“機率0%”，或“100%”。一事件之發生機率，不論多小，只要不是0，就可能發生。而發生機率不論多大，只要不是1，便可能不發生。

投擲一個公正的銅枚100次，正面出現的相對頻率會接近0.5(大數法則)，總共出現之正面數則在50附近，不見得會剛好等於50。至於會不會正面數偏離50很遠？極端一點，譬如說連得100個正面？此機率為0.5的100次方，按一下計算機，小數點後面連續30個0才出現7。除了小說裡，這種事件，顯然很少能見到。只是機率雖微乎其微，卻仍是正的。事實上，小機率事件，就是常被人們稱為奇蹟，或說不可能的事件。諸位看過湯姆克魯斯(Tom Cruise)主演的“不可能的任務”(Mission：Impossible)嗎？第一集是1996年上映，至2006年已經有三集。每一集中之不可能的任務，最後都完成了。要知“不可能”是口語，並非機率裡的術語。原先你以為不可能的事情，只要觀測次數夠多，就不難發生。有人稱此為**巨數法則**(law of truly large numbers)。當小機率遇上大樣本，其發生就不太令人驚訝了。

毛澤東在“和郭沫若同志”的“滿江紅”那首詞中，有句“一萬年太久，只爭朝夕”。政治人物多半只在乎當下。對於機率，我們則應不爭一時而爭千秋。對機率的品頭論足，說三道四，並不能只看少數幾次的結果。一旦觀測次數夠多後，機率的威力就顯現出來，這時僥倖便只能靠邊站了。有一個銅板，出現正面機率為0.6。投擲若干次，那一面出現較多次便贏。請問你要選那一面？雖正面比反面出現的機率高，但假設你不信邪，偏要選反面。則若僅投擲1次，你有0.4的機率贏，不算太差。投擲10次時，你約有0.166的機率贏，也還可以。但投擲100次時，你贏的機率就僅

約0.016了。至於投擲1,000次時，你贏的機率，是小數點後面連續10個0才出現4，已極接近0了。經年累月下來，我們要做許多許多的決策，少鐵齒，依機率行事，還是較明智的。

幾年前有本“與天為敵”(英文書名是Against the Gods)，曾一時造成洛陽紙貴。此書宣稱運用機率及統計方法，以控制風險，因而能“趨吉避凶，天威不再難測，人類的未來得以擺脫諸神恣意的作弄”。有些勵志的書裡則說“雙手萬能，人定勝天”。大家也都知道，一直有人留連賭場，有人沈迷彩券，追求明牌，想要一夕致富。我們的看法是，趨吉避凶，與天為敵，及人定勝天等，都是奢望。與天為友才是對的。而對於賭，就以樂透彩為例。號稱公益彩券，發行銀行拿出來當獎金的，約是總投注金額的56%。你想豪賭，拿出10萬元投入，中的獎金繼續投入。每週開兩期，你知道1個月也就是8期後，會成為多少錢嗎？隨機現象，當然不一定，說不定你財星高照，8期皆中頭獎。只是不難理解，你的錢將大約以每期0.56的比例下降。8期後約成為10萬元乘上0.56的8次方，差不多是967元，下降速度驚人。投資彩券，將會如紅樓夢四十七回裡鳳姐所說“這一吊錢，頑不了半個時辰，那裡頭的錢就招手叫他了。”沒辦法，機率是這樣告訴我們。十年才能贏得青樓薄倖名。但想跟樂透彩拼的人，要贏得公益名，所花的時間，可是超乎想像的短。

你可以拼經濟，也可以拼政治。但在隨機世界中，卻要相信機率，而不要挑戰機率。

4 合理估計

從前有一個賣油條的小孩，他一向把賣得的錢，都放在盛油條的籃子裡。某日由於尿急，於是把籃子放在一塊大石頭上，解放去也。過一會兒回來，晴天霹靂，籃子裡的錢都不見了。他哭著跑去告訴縣官。縣官聽了後，叫人把石頭抬來審問。雖一再恫嚇，石頭一句話也不說。縣官氣了，叫人拿棍子來打石頭。只是即使打到棍子斷了，石頭仍不說話。一旁看熱鬧的人都笑了起來。縣官更生氣，罰圍觀者每人拿兩個銅錢，扔進一個

盛滿水的盆子裡。突然，縣官指著一個人說“偷錢的人就是你。”那人大呼冤枉，眾人也不解。縣官解釋說，“那小孩是賣油條的，他的錢上都沾著油。別人的錢扔進水裡都沒有油浮上來，只有這個人扔錢進水後，有油浮上來，可見錢是這人偷的。”那人俯首認罪，眾人皆心服。

憑口袋裡的錢有油，就認定他偷了賣油條小孩的錢？如果有人收到賣油條者找錢，不也就沾著油嗎？這則民間故事，有不同的版本，如“包公審錢案”。在舊約聖經列王紀上第3章，當二婦人爭奪小孩不下時，**所羅門王**(King Solomon)叫人拿刀來將小孩劈成兩半，一人一半。二婦人反應不同，一位就放棄爭取了，一位則贊成將小孩劈了。所羅門王遂將小孩判給放棄者，因他認為生母是不忍心讓小孩成為兩半的。這種包公、所羅門王式的智慧，與教室玻璃破了，老師先從平常最調皮者問起的原理類似：當從幾個可能性裡做挑選時，優先挑最可能的情況。會不會誤判？當然也是會的。清朝劉鶚的“老殘遊記”中，有一“剛弼斷案”的故事就是例子。你以為很可能會怎麼樣，有時卻不一定。要有隨機的概念，做決策時，不能剛愎自用。聖經列王紀上第4章，說“所羅門的智慧超過東方人和埃及人的一切智慧”。這種人們在做選擇時，常採用的所羅門王式的思維，在統計裡就發展出**最大概似法**(method of maximum likelihood)—依發生機率之最大者來決定估計值。此估計法有很多好的性質，常也能得到不錯的估計量。

美國NBA職業籃球賽，各球隊互有勝負，很難說那一球隊才是最強。在常規比賽裡，每支球隊要賽82場，各區勝率最高的8隊可打季後賽。所謂勝率，就是贏的場次除以比賽場次。為了維持比賽之可看性，NBA有一套選秀機制，使各隊實力不會很懸殊。有時全季排名第一者，勝率還不到6成。以一個球季多場比賽後的勝率，決定誰是今年較強者，得以參加季後賽，是職業球賽常採的作法。諸如估計某項手術的成功機率，估計生三胞胎的的機率，也是常採用這種以“**相對頻率**”來估計的想法。因而就發展出**動差法**(method of moments)。為統計中一重要且常是不錯的估計法。

不論生活上及科學上，人們常在做估計。估計這綠燈過得去嗎？估計追上心儀女孩子的機率。估計這樣的成績，能不能上台北大學統計學

系。估計中國大陸野生熊貓的數量等。比較嚴肅一點的，估計台灣明年經濟成長率，估計20年後台灣每年新生兒人數。百家爭鳴，可以有各種估計法。有道理的估計法，多半有其優點，各有其適用的時機。就如武俠小說裡，少見有那一種武功是天下無敵的。而一個停止的時鐘，每天都有比一天只慢千分之1秒的鐘還準的時刻。統計裡的估計量，是有評比方法的。但所謂優劣，乃依評比指標而定。各擅勝場，不會有那一估計法，是永遠最佳的。我們再給一種估計法。有時依據過去經驗，或主觀上的認定，會有一些事先的看法。比方說，一表人才，又彬彬有禮的人，我們對他的初步印象，很可能會不錯。有較好的外表，總是無言的推薦。學歷不錯者，常也會吸引人，血統仍是頗具參考價值的。但相處一段時間後，只見做事顛三倒四，對其評價往往就會做些調整。又如，對銅板出現正面的機率，我們覺得銅板是中央銀行製造的，不致於太偏頗，應就在0.5的附近，如0.49至0.51間。至於這區間中的點，可能性就認為都差不多。換句話說，對於正面出現的機率，你覺得有一事先的分佈。這分佈，是所謂在區間0.49至0.51的**均勻分佈**(uniform distribution)。但實際投擲多次後，如果正面出現的相對頻率是0.68，這時你總不致於還堅持出現正面的機率應在0.49至0.51均勻分佈。一意孤行畢竟是不好的，雖事先有主見(即使主觀，有時也仍會參考過去的資料)，但依觀測後的結果，調整原先的看法，這也是人們常有的一種思維方式。在統計裡，就發展出**貝氏法**(Bayesian approach)。

在**論語**先進篇中，孔子要幾個弟子分別講其抱負，其中出現了一些數字，像是“方六七十，如五六十”，又如“冠者五六人，童子六七人”。我們常想知道明天溫度為何？氣象局提供一個很明確的預報—27度。病人能活多久？醫生說3個月。但一天中的溫度，從早到晚總是有差異的。病人也少有因醫生的直斷，就恰好3個月後死亡。因此有時我們覺得給個範圍，如孔子那幾個弟子的作法，可能是更清楚的描述。統計裡便發展出**信賴區間**(confidence interval)估計法。

以某候選人的支持度為例。一個常見的講法是“支持度是0.35，在百分之九十五的**信心水準**下，抽樣誤差在正負三個百分點以內”。或者說有百

分之九十五的信心水準，支持度介於0.32至0.38間。或說區間(0.32,0.38)是支持度之95%信賴區間。信心水準又稱**信賴係數**，而95%就是0.95，其實是一機率值。只是為什麼不稱機率而稱信心水準，到底有什麼信心？究竟在信賴什麼？

政治人物慣說“我對XXX是百分之百信賴”。實則彼此貌合神離，勾心鬥角，毫無互信可言。對於估計一銅板出現正面的機率 p ，**點估計**是以一個值(稱做**估計量**)來估計。雖然明確，但估計值要恰好等於 p ，其機率常是微乎其微，甚至可說是0。因此我們對以單一的值來估計 p ，信心自然是不太夠。區間估計是以一隨機區間來估計 p 。此區間表 p 之可能落在的範圍。信心水準就是指 p 會落在此區間的機率。就估計的觀點，一區間當然較一點更可信賴。我們更有信心此區間會包含 p 。因此稱此區間為信賴區間，有其道理。只是一旦取完樣本，所得之信賴區間，為一確定的區間。這時該固定區間，或包含 p ，或不包含 p 。一翻兩瞪眼，已不能再說會包含 p 的機率，是原先的諸如0.95，或0.99了。這是對這些95%，或99%，避開機率一詞，而稱為信心水準的原因之一。

給出一估計之明確的信心水準，而不像有些人，常掛口中，可信度不高的“百分之百信賴”，或“完全有信心”，反而使此區間更可信賴。至於有多信賴？信心水準就是給出信賴程度。

從估計的角度，在同一信心水準下，信賴區間愈狹窄愈好，這表示估計愈精準。只是通常這要增加樣本數才行，得有成本的考量。又大家是否好奇，為什麼不就給百分之百的信賴區間？不要有誤差不是很好嗎？只是若要零風險，這種區間往往是很長的，反而失去參考價值。此有如百分之百追得上的女孩子，男生的興趣通常可能不太高。百分之百成功的投資，不是你沒興趣，就是輪不到你。在隨機世界中，變數太多，一切的估計，都不過是個參考。追求太高的信心水準，有時得付出不小的代價，不見得值得。

5 無罪推定

人們常求公平或公正。以簡單的兩人分蛋糕為例，若雙方皆不願拿得比較小，那有什麼好方法來分？你切我選，應是一個令兩人都不覺得吃虧的辦法。最好是連由誰切，都以抽籤的方式。以免選方感覺他所得大於一半，而切方感覺他所得只有一半。

專制時代，寧可錯殺一百，不能錯放一個。民主時代，即使沒有包青天、所羅門王，法庭上如何設計出一套，讓人覺得較公正的判決機制？宋朝歐陽修，在追述其父母生前言行事蹟的“瀧岡阡表”一文中，提及幼時母親曾告訴他，其父生前為官批文時，對於死囚，一向的做法是“求其生而不得，則死者與我皆無恨也”。以往檢察官，若認為法官未窮盡調查之途徑，便判被告無罪，會不服而提起上訴。那是因我國最高法院，於民國25年立下“有罪推定原則”之判例。經過60餘年，終於在民國92年，在刑事訴訟法裡，修正為“**無罪推定原則**”。從而檢察官須善盡舉證責任，證明被告有罪，俾推翻無罪之推定。

法官之判決，採無罪推定原則是較有道理的。即先假設被告無罪，然後看在無罪之下，為何會有這些啓人疑竇的事證產生？若被告無法合理交待，只好判他有罪了。但這不就是一千年前，歐陽修的父親(歐陽修生於西元1007年，4歲喪父)所秉持的盡力為死囚“求其生”的精神嗎？無罪推定原則，便類似你切我選，屬於能令檢察官與被告，皆感到較公正的一種判決法。

無罪推定原則的影響是較正面的。假設有一被起訴者，最後因法官認為證據不足而判其無罪。如果他真無辜，那此為一正確的判決；如果他其實有罪，但被釋放後洗心革面，放下屠刀，再也不犯罪，那也很好；如果他因心存僥倖或其他原因，再度犯罪，則夜路走多後，不見得每次都有那麼好的運氣，終有因證據充分，而被繩之以法的一天。但若採有罪推定原則，被起訴者較容易被判有罪，將造成較多的冤獄。而一旦執行刑罰，日後如果發現原來是誤判，就不易彌補了。

某次考試有20道選擇題，每道題有4個選項。我們來看下述二情境：

1. 老師對學生說如果你們二位沒作弊, 怎會有15題對的一樣?
2. 如果你們二位沒作弊, 怎會有15題錯的一樣?

實際的情況是, 20題中, 有15題較容易, 大部分的學生都能選對答案, 兩人這15題都答對, 是一件很尋常的事, 所以不能據此推定他們作弊。但要錯同樣的15題, 且選項也都相同, 這就很不尋常了。這兩個學生要說服老師他們沒作弊, 大約很難。

1933年, 波蘭人奈曼(Jerzy Neyman, 1894-1981), 及英國人皮爾生(Egon S. Pearson, 1895-1980), 給出著名的奈曼-皮爾生引理(Neyman-Pearson lemma), 奠定了統計學裡的無罪推定原則。較我國刑事訴訟法, 早了整整70年。奈曼-皮爾生提出了一套假設檢定(hypothesis testing)的架構。在其架構中, 有一虛無假設(null hypothesis), 常以 H_0 表之, 及一對立假設(alternative hypothesis), 常以 H_a 表之。虛無假設通常表現況, 或我們傾向不相信者, 而對立假設則表我們傾向相信者。以樂透彩開獎為例。42個號碼, 每次開出6個頭獎號碼, 如果號碼是隨機產生, 每個號碼出現的機率應為 $1/7$ 。有人懷疑1號比較容易出現, 於是來做個檢定。先將虛無假設取為“1號開出的機率是 $1/7$ ”, 至於對立假設則取為“1號開出的機率大於 $1/7$ ”。此外, 依不同的情況, 虛無假設也可以是: 喝綠茶不能減肥, 某模特兒沒有服用毒品, 新教材與舊教材效果一樣等。而對立假設則可以分別為: 喝綠茶能減肥, 某模特兒有服用毒品, 新教材比舊教材效果好等。

大家不難想通何以 H_0 被稱為虛無假設。警方因接獲密報, 懷疑某模特兒服用毒品, 才找她來檢驗, 媒體也大肆報導。若大費周章後, 宣佈該模特兒無辜, 則輿論難免會同情該模特兒受了無妄之災。幾次這種事件下來, 警方可能會被認為擾民了。再以法庭為例。雖說依無罪推定原則, 但法官若認定起訴者無罪, 檢調單位其實是灰頭土臉的。所以 H_0 被接受, 往往表做虛功, 回到原點, 只能接受一空的假設。

英文中的假設hypothesis, 是由古希臘文 hypotithenai演變而來, 科學上的假說(或稱假設學說)也是這個字。大家在中學的化學課程裡學過, 義大利化學家亞佛加羅德(Amedeo Avogadro, 1777-1856), 在1811年

提出，於相同的物理條件下，相同體積的氣體，含有相同數目的原子。但此觀點未被當時的科學家所接受，因此被稱為亞佛加羅德假說(Avogadro's Hypothesis)。其後被證實為真，就改稱定律(Avogadro's law，或Avogadro's Principle)。在數學裡，我們常在證明一命題是真或偽。但在隨機世界中，很多現象都只能視為假設，就看接受那一個。接受不表示就完全相信該假設為真，拒絕也不表該假設為偽。就如女孩子找對象，最後決定嫁誰，說不定也只是認為該男子可以接受而已。而一對象之接受與否，其實乃視標準之高低。統計裡的假設，經檢定後，不論接受那一個假設，都無法讓該假設成為定律，假設永遠是假設。法官判定常也是類似。所謂“司法還我清白”，都是被判無罪者自己說的。你想探索真相嗎？真相常難以大白。此正如“周公恐懼流言日，王莽謙恭下士時，若使當時身便死，一生真偽有誰知？”又如日本那部著名的電影“羅生門”(1951)，同一事件的4位目擊者，每個人的陳述都不同，電影最後也沒告訴觀眾，究竟誰說的才是真話。那4人說不定也不是存心說謊，而是每個人觀點及判斷不同。金庸的“雪山飛狐”一書中，也有類似的羅生門情節。因此通過檢定的假設，只是表在現況下，比較該接受的一個選擇。接受一假設後，常要輔以其他科學方法，以進一步探討。

在一統計檢定裡，不論拒絕或接受 H_0 ，都可能犯錯。這其中有兩種錯誤，第一型是 H_0 為真卻拒絕 H_0 ，第二型是 H_a 為真卻接受 H_0 。最理想的，當然是兩型錯誤機率皆為0，但通常不會有這種情況。虛無假設是要較被保護的。想想若明明各號碼出現機率相同，卻宣佈1號較易出可見，造成的影響將有多大？明明沒服毒，卻被判定服毒，有多冤枉？奈曼-皮爾生的作法是，先給定一個所能容忍的第一型錯誤之機率，這當然是一個比較小的值。依據此機率，決定得到什麼樣的觀測值時拒絕 H_0 。至於此機率該多小？乃視犯第一型錯誤的後果有多嚴重。如果 H_0 是嫌犯沒殺人， H_a 是嫌犯殺人，那所能容忍之第一型錯誤的機率，當然要很小才行。如果對於某食品， H_0 是不會致癌， H_a 是會致癌，這時就不宜過度保護廠商，而忽視消費者之健康。因此該容忍之第一型錯誤的機率，就不用太小了。

一觀測值若使 H_0 被拒絕，此觀測結果便稱為**顯著的**。要注意的是，顯

著與否，並非依觀測值的大小，而是依發生機率之大小。發生機率大的事件，為一尋常事件，不為顯著事件。所謂尋常事件，就是看到後不用大驚小怪。發生機率小的事件，才是顯著事件，才會導至 H_0 被拒絕。做統計檢定時，與媒體人一般，要睜大眼睛留意是否有顯著事件發生。

由於無罪推定原則，兩人之間若有爭議，各說各話，僵持不下，則除非證據夠，否則提起控訴的一方，不見得有利。我們給一實例。 A 公開宣稱他當年所公佈的那卷XXX緋聞錄音帶，是 B 所提供的。 B 獲知後，控告 A 誹謗。地檢署經調查，認為“並無積極證據證明 A 所說是虛構”，因此對 A 做出不起訴處分。雖然檢方並未說此即認定錄音帶是 B 交給 A 的，但媒體刊登的新聞標題，卻為“XXX緋聞錄音帶，檢方認定 B 給的”。同樣的道理，除非跡象顯示假設 H_0 很可疑，否則不會輕易啟動一個擬推翻 H_0 的統計檢定。

前面說過，顯著與否，乃依發生機率之大小，而非觀測值之大小。底下給一例。

某次總統選舉，共有 A 、 B 兩組候選人，競爭激烈。投票結果 A 組得票率50.114%，僅比半數的50%略多一些。落後的 B 組很不服氣，覺得這麼小的差異，在統計上是沒有意義的。究竟如此小的領先，對區分兩組候選人得票有差，夠不夠顯著，其實乃與投票人數有關。取 H_0 為 A 、 B 兩組沒有差異， H_a 為 A 組獲得之支持度大於2分之1。先假定投票人數為10萬人。則在 H_0 為真之下，會得到至少有此領先之機率，約為0.2355，接近4分之1，可說很容易發生。因此這是一相當尋常的事件，不易據此以拒絕 H_0 。換句話說，投票結果顯示 A 、 B 二組候選人所獲得之支持度，並沒什麼差異。亦即 B 組的支持者不服氣尚有點道理。其次假定投票人數為1百萬人，則在 H_0 為真之下，會發生至少有此領先之機率，約為0.0113，算是很小了， B 組最好就承認敗選。最後假定投票人數為1千3百萬人。則在 H_0 為真之下，會得到至少這樣領先的機率，約為1.03乘上10的負16次方，即小數點後連續15個0才出現1，可說微乎其微。因此看到此一投票結果，還認同兩組候選人之支持度沒有差異的假設，可說是毫無道理的。換句話說，在投票人數高達1千3百萬人之下， A 組得到這樣的領先，是很顯著的。足以拒

絕 A 、 B 兩組沒有差異之 H_0 的假設。

最後必須一提的是，一般而言，一項統計檢定，即使再嚴格，也無法決定因果關係。民國97年各高中推薦在校成績前1%的學生參加“繁星計畫”，再經大學挑選，以進入大學就讀。位於台中縣的常春藤高中，有254名高三學生，報上說“理論上，可有3位學生獲推薦，結果該校共推薦31人，且最後有11人錄取，明顯高過正常比例。”這個“明顯高過正常比例”，用統計語言來說，就是一“顯著事件”。但統計的功能只能到此，該校成績究竟是否作假，必須調出學生原始成績才知。

6 紙上談兵

朝野猶誇紙上兵。紙上談兵本來是比喻有些人只會空談，不顧實際。戰國時代，趙國的趙括，其父趙奢為一代名將。雖家學淵源，又自幼熟讀兵書，能言善辯，談起軍事頭頭是道。趙奢卻不認為趙括可領軍，因“兵，死地也，而括易言之。”知子莫若父，趙奢認為兒子光會紙上談兵，不知用兵之難。藺相如也曾對趙王說“括徒能讀其父書傳，不知合變也”。只是後來趙王仍以趙括代廉頗為將，趙括也果真讓趙國大敗，生平事蹟成為“紙上談兵”成語之由來。

沒有資料則難以做決策。在統計裡，資料常得仰賴取樣。理論裡的樣本數，說要多大就多大。想要趨近至無限大，無限多個樣本彷彿就立即出現在眼前（這正是紙上談兵）。只是不論醫學上，或商場上，有時樣本並不易獲得。而諸如做民調時，樣本常須在短時間內取得。想要增加樣本數，往往亦有實際困難。除了樣本不易大量取得，樣本的偏差，更會造成決策的失準。例如，一種新藥被開發出來，要找些人來做實驗，能以自願的方式嗎？不行，因這樣可能來的比較多是病入膏肓者，打算死馬當活馬醫。所以通常要將接受實驗者分為兩組，一組是**處理組**(treatment group)，一組是**控制組**(control group)。處理組服用新藥，控制組則不服用新藥。接受實驗者該屬於何組，採隨機的方式，以消除偏差。若無控制組，如何判定服用此藥後確實有效？藥效又多大？又接受實驗者者，及參與的醫護人員，

都不能知道誰是那一組。免得因一些心理因素，造成數據的偏差，因而影響實驗結果。要知通常醫護人員對屬於處理組者，可能不自覺會對他更關懷。而大家也都知道，病人若覺特別被照顧，生存意志可能會大些。因此屬於控制組者，也要服用一種外觀與新藥一樣，但毫無任何作用的安慰劑(placebo)。

不論醫學上的實驗，或選舉的民調，都不乏雖大量取樣，結果卻完全失敗的例子。如何以較經濟的方式，取得較客觀的樣本，而能達到同樣(甚至更佳)的準確度，就必須先紙上談兵，妥善規畫了。不但選擇適當的統計方法很重要，統計學裡因此也發展出**實驗設計**(experimental design)。亦即做實驗以取樣前，要先設計良好的取樣程序。另外，因實際數據之不易得，有時得先模擬產生數據，紙上談兵一番。看看擬採用的方法到底好不好，有如軍事上的兵棋推演。不要以為虛擬的數據，都是有如海市蜃樓，不過是幻影。事實上，只要模擬的好，對實務常有很大的幫助。清末黃花崗七十二烈士之一的林覺民，在“與妻訣別書”中，寫不盡對愛妻的不捨。最後說“紙短情長，所未盡者尚有幾萬千，汝可以模擬得之。”此愛綿綿無絕期，要愛妻依據信中所述的內容，自行揣摩想像其他的萬千字句。連情意都可模擬，數據當然更能模擬了。林覺民真可說是現代統計模擬的先驅。

統計由於並非只是在紙上談兵，純做理論的探討。而當實際應用時，就會有各種成本上的考量，因而就更需要並非空談之紙上談兵了。

結語

1985年11月，有位美國學者泰勒(Gary Taylor, 1953-)在英國牛津大學(University of Oxford)的一圖書館，找到一首詩，就稱為泰勒詩(Taylor poem)好了。英美研究**莎士比亞**(William Shakespeare, 1564-1616)的學者們，為了此詩是否為莎士比亞所作，而大打筆戰。不少專家認為這首泰勒詩，不論用字遣詞，與韻味風格，都迥異於莎士比亞其他作品。筆戰兩個月後，1986年1月24日出版的Science雜誌，刊登一篇“莎士比亞的新詩-向統計學禮讚”(Shakespeare's new poem: an ode to statistics)，介紹兩位統計學者Efron與Thisted，如何以統計方法，鑑定這首泰勒詩，是否為莎士比亞

所作。

Efron與Thisted的想法是這樣的：每個人各有其用字習慣，特別是對於罕用字，每個作者使用的習慣，差異可能更大。莎士比亞已知的總作品中，共有884,647個字，其中有31,534個相異字。這些相異字中，有14,376個字從頭到尾只出現1次，有4,343個字只出現2次。出現幾次的字，都被計算出來。那些在總作品中，出現頻率較低的，就是莎士比亞的罕用字。依據這些資料，假設這首共429個字的泰勒詩，為莎士比亞所寫，他們估計會有幾個字，在總作品中從未出現(也就是新字)，只出現1次，2次，…，一直到曾出現99次，都給出估計值。實際情況與估計非常吻合。這樣做還不夠，會不會當時代的詩人，用字習慣都差不多？兩人又找了三位約略與莎士比亞同時代的詩人，各取其一首詩，及另取四首莎士比亞的詩，與這首泰勒詩做比較。經過3種統計檢定，發現對前三首，若假設為莎士比亞的作品，罕用字出現次數之實際值，與估計值皆不吻合。而所挑選的四首莎士比亞的詩，雖偶有不合，但總的來說，是可接受的。Efron及Thisted說，他們的分析，並無法證明泰勒詩為莎士比亞所寫，但在罕用字之使用情況，如此與莎士比亞的總作品吻合，確實令人驚訝。

一場文學上的爭論，經統計學家發聲後，迅速平息，難怪要向統計學禮讚了。運用統計方法來做決策，反映的是一種客觀及合理的思維。要知與其主觀的爭論風格相同否，彼此都無法說服對方，還不如以客觀的統計方法來判定。但如何才算已經夠客觀？除了只檢驗泰勒詩外，Efron及Thisted亦拿幾位與莎士比亞同時代的詩人來比較，這樣就更保險了。免得萬一莎士比亞那個時期的詩人，有如時尚般，罕用字之使用習慣類似，則此檢定就沒有什麼參考價值了。欲進一步了解此事之來龍去脈，可參考黃文璋(1999)第十七章“莎士比亞新詩真偽之鑑定”，相關文獻也可在該文找到。

但統計的判定，是否一出手，就令人臣服呢？

克萊門斯（Roger Clemens, 1962—），外號“火箭人”，是一位在美國職棒大聯盟，活躍二十餘年的投手，被認為是大聯盟史上最偉大的投手之一。至今他一共獲得七座賽揚獎(Cy Young Award)，為歷來得到該獎最

多的大聯盟投手。此獎設立於1956年，起初每年選出一位。自1967年起，兩個聯盟每年各選出一位。能得到一次都不容易了，何況七次？克萊門斯在賽揚獎的成就，大約很難被超越。2003年球季結束後，克萊門斯宣佈退休，但隔年便改變心意，重回大聯盟。且在當年以42歲“高齡”，再度拿下賽揚獎。他自2007年起效力於紐約洋基隊(New York Yankees)，成為王建民隊友。只是由於他的前訓練員麥克納米(Brian McNamee)，指控他在職業生涯的後期，服用類固醇(steroid)及生長激素(human growth hormone)，於是美國國會進行調查。克萊門斯當然要盡力維護其聲譽，一世英名豈能毀於一旦？他聘請的專家為他整理出一份長達45頁的報告，包含38個圖表，想為他洗清冤屈，證明他是老而彌堅，而非依靠禁藥之助。只是一份報告的價值，當然不是以厚度來度量。曾有得諾貝爾經濟獎的論文，只有二十多頁。4位美國賓州大學(University of Pennsylvania)華頓學院(Wharton School)的統計以及經濟教授，仔細檢視後，認為這份報告，對讓人相信克萊門斯的無辜，並無說服力。

究竟克萊門斯在職業生涯的後期，仍虎虎生威，是否為極不尋常(highly unusual)的事件？如果是的話，則他的前訓練員所宣稱的，自1998年起，替克萊門斯注射生長激素，就可能為真。為了化解此疑慮，在克萊門斯所自行提出的這份報告中，將他自己，與在1993年46歲時退休的名投手萊恩(Nolan Ryan)進行對比。資料顯示，克萊門斯的表現看起來沒什麼不尋常，兩人都是在40多歲時達到顛峰。再將克萊門斯與另外兩位同時代的投手詹森(Randy Johnson)，及席林(Curt Schilling)相比，結果也類似。一切似乎無懈可擊。數據會說話，該還克萊門斯清白了吧！

這4位專家可不買這些數據的帳。他們指出，報告中僅是將克萊門斯與那些“在職業生涯第二階段獲得成功的投手”相比較，而不是與所有跟克萊門斯一樣，在菜鳥階段就揚名立萬的投手相比。在統計學裡，稱此為“選擇偏差”(selection bias)。一旦比較的對象精心挑選，克萊門斯的數據，沒有顯得不正常就不稀奇了。這4位教授的文章，發表在銷路廣大的“紐約時報”(New York Times, 2008年2月10日)，題目是“支持克萊門斯的報告乃精心挑選真相”(Report Backing Clemens Chooses Its Facts

Carefully)。口說無憑，4位專家另起爐灶，分析自1968年起，大聯盟31位優異的投手之生涯投球數據。相對於這較大的比較群，克萊門斯的數據，就顯得異常了。一般投手的表現，都是初期逐漸成長，在30歲左右達到高峰。之後大約自35歲起，便廉頗老矣，表現就逐漸走下坡。但克萊門斯卻不然，他的表現在快達30歲時下降，而在35歲至40歲間又呈現成長，老幹贏新枝，顯得極為不尋常。4位專家的分析，自然無法證明克萊門斯，在球員生涯之後期服用禁藥。統計並非萬能，不見得回回能一言而為天下法。但他們使克萊門斯自行聘請的專家，所做之報告黯然失色，並使克萊門斯仍無法自服用禁藥的疑雲風波脫身。統計的威力，仍再度顯現。

本文一開始說“有三種謊言：謊言，可惡的謊言，及統計”。統計為何被當做謊言呢？可惡的謊言已夠可惡了，難道統計是更可惡的謊言？上二例已足以說明。統計的結論要有價值，要被人採信，其中涉及的每一程序，從設計，取樣到分析，都要竭盡所能，儘量客觀。統計學家會犯錯，因所有保證都是機率式的，並附帶一定的犯錯機率。決策若不願犯錯，後果不見得就好。法庭若不願犯錯，可能使大部分明明有罪的人，都被釋放。但機率理論也告訴我們，如果統計分析是遵循該有的程序，則長期下來，犯錯次數的比例，差不多就是所設定的犯錯機率，乃可容忍。但分析過程中，若有意或無意的偏差，則即使工程再浩大，得到的結論，不但無法取信真正的專家，被當成謊言不說，有時還給自己製造出極不利的後果。

統計，正如我們的思維，客觀至上，否則便是自欺欺人。反之我們的思維若是統計式的，便是極客觀的。

參考文獻

1. 黃文璋(1999). 數學欣賞。華泰文化事業公司,台北。
2. Bradlow Eric, Jensen Shane, Wolfers Justin, and Wyner Adi (2008). Report Backing Clemens Chooses Its Facts Carefully. New York Times. February 10, 2008.