統計下凡

黃文璋

2022 年 10 月

目　錄

1 球賽裡的機率

眾所周知的 NBA(National Basketball Association)，乃北美(包括美國及加拿大)男子職業籃球聯盟，共有 30 支球隊。依球隊所在城市的地理位置，分東西兩區，每區各有 15 支球隊。每年依例行賽勝率排名，東西區各取 8 支球隊進入季後賽(因疫情的關係，自 2020 年起，各區例行賽的第 7 至 10 名要加打附加賽，以決定那兩隊成為能打季後賽的第 7、8 名)。季後賽依強對弱的原則，且採 7 戰 4 勝淘汰制。即第 1 輪進行每區排序第 1 對第 8，第 2 對第 7，第 3 對第 6，第 4 對第 5 的系列賽，先獲 4 勝者晉級。每區進入第 2 輪的 4 支球隊，由第 1 輪第 1 對第 8 的勝隊，對上第 4 對第 5 的勝隊;第 2 對第 7 的勝隊，對上第 3 對第 6 的勝隊，兩支勝隊晉級第 3 輪。第 3 輪的 7 戰 4 勝後，產生的東、西區冠軍再對打，先贏 4 場的球隊，便是年度總冠軍。

NBA 的每支球隊，每年均有兩個目標，第一是進入季後賽，第二是得到總冠軍。能進季後賽，大致表這季的成績及格；而獲得年度總冠軍，對球隊及球員，甚至球隊所屬城市之居民，皆為莫大的榮耀。整個季後賽可說高潮迭起，場場扣人心弦、勝負令人關切。若毫無資訊，則“相同可能性”的概念，當然能派上用場。即兩支球隊交手，可假設兩隊獲勝的機率皆相同，即各為 1/2。至於 30 隊，則假設每隊得年度總冠軍之機率皆相同，即各為 1/30。只是 NBA 歷年來各種數據收集極為齊全，一場比賽才剛結束，球隊或某位球員

破了什麼紀錄，立刻公佈。齊全的數據，使在 NBA 的預測勝負時，相同可能性的想法，很少能用得上。

競爭激烈下，風光難以維持太久。2020 年季後賽通過第 1 輪的 8 支球隊，2021 年僅剩 3 隊闖過第 1 輪；而 2020 年，爭奪總冠軍賽的湖人與熱火兩隊，更在 2020 年首輪便皆被淘汰出局。除了球是圓的外，基於各種因素，如球員可能失常、主場與客場有別(球迷的掌聲與噓聲，相當程度地影響球員的心理)、球員可能因傷下場或缺賽(短者缺幾場，長者整個球季都無法上場)，再加上有時教練會讓主力球員輪休，因而球隊的表現，有很多不可測的因素。以 2021 年為例，NBA 的例行賽，每隊皆打 72 場(通常打 82 場，此季因疫情減少)，30 支球隊中，勝率最高的爵士隊輸了 20 場(贏 52 場，勝率約 0.722)，輸的球隊中，不乏未進入季後賽者。至於勝率最低的火箭隊，仍贏了 17 場(勝率約 0.236)，贏的球隊中，包括幾支強隊。可見再強的隊伍也沒有打不敗的，且再弱的隊伍也不見得完全不堪一擊。

不妨以簡單的投擲銅板為例。假設銅板 A 出現正面的機率為 0.722，銅板 B 出現正面的機率為 0.236。今各投擲 1 次，該如何預測出現的結果？預測銅板 A 出現正面，且銅板 B 出現反面，顯然極合理。但若有人鐵齒，預測銅板 A 出現反面呢？他有 0.278 的機率預測正確。此機率其實並非太小，約每 3、4 次便會有 1 次正確。現在來模仿 NBA 季後賽的晉級方式。投擲銅板 A，要賭正面或反面先出現 4 次？不必想，便知該選正面。而且，由於賭對有 4 正、4 正 1 反、4 正 2

反，及 4 正 3 反等 4 種可能，且後 3 者的最後 1 次投擲須出現正面，利用排列組合，可求出贏的機率約為 0.90068。超過 9 成的機率，算是相當高。但仍有約 1 成的機率會陰溝裡翻船。沒辦法，隨機現象就是如此。

球隊比賽與投擲銅板的情況當然大不相同。對於銅板，可耐操耐煩地讓人一再投擲，因而每次投擲出現正面的機率相同，且各次投擲間出現的結果相互獨立之假設，挑戰的人大概不會太多。至於球賽，不要說各場比賽對手不盡相同，且上場球員可能有異，就算對手及上場球員皆相同，球員的心理及體能等，也難維持不變。因而依前幾場球隊之表現及球員之近況，教練的戰術可能調整。即使球賽進行中，也隨時可能依戰況做調度。故對一特定的球隊，各場比賽獲勝的機率，難以假設皆相同，不同比賽間的結果，也難視為相互獨立。對一有兩個結果之隨機現象(如比賽有輸、贏，議題有支持、不支持等)，常以簡單的投擲銅板來模擬，已發展出不少令人讚嘆的理論，並廣泛應用。但實務上，很多情況與投擲銅板當然大異其趣。因而若引用投擲銅板的理論，誤差不要說難免，若不謹慎，得到極荒謬的結論乃不足為奇。

ESPN(Entertainment Sports Programming Network)是美國一個 24 小時專播體育節目的有線電視聯播網。2021 年季後賽開打前，ESPN 所找來 18 個專家之分析預測中，有 17 個認為西區第 4 種子快艇隊在 6 場內(即 4：0、4：1，或 4：2)，便會淘汰第 5 種子獨行俠隊，順利打進西區季後賽第 2 輪。一般認為，不論從球隊與球員的過往戰績，以及球隊的

近況(所謂利用大數據)，此預測相當合理。那我們可以說，快艇淘汰獨行俠的機率為 17/18(即約 0.9444)嗎？並無不可。正如若投擲一銅板 18 次，出現 17 次正面 1 次反面，有人遂以正面出現數之“相對頻率”17/18，來估計銅板出現正面的機率。

投擲銅板是用同一個，這群專家卻是不同的人，有人指出此疑點。是沒錯，但比賽前，兩隊究竟何者會晉級，本來就只有天曉得，17/18 的機率，乃基於 18 個專家的看法，這難道不是一還算客觀的估計嗎？此作法並非不合理，總比 1 個人瞎猜精準。那有其他估計快艇會進入第 2 輪的機率之預測方式嗎？也是有的。例如，在例行賽中，兩隊共交手 3 次，獨行俠取得 2 勝 1 敗的對戰優勢。獨行俠例行賽戰績雖不如快艇，或許打快艇卻特別有一套。若以 1/3 做為兩隊每次交手，快艇獲勝之機率，則快艇能晉級第 2 輪之機率，將由相當高的約 0.9444 降至僅 379/2187，即約為 0.1733。只是這樣的估計，很可能被會挑剔，因只憑交手少少的 3 次來估計，很難令人覺得夠精準。若兩隊曾對戰 10 次，此時以勝負結果來估計，準確度會被認為較高。

另外，說不定有人以為，進入季後賽便一切重來，他主觀地認定快艇與獨行俠能過關的機率相同，即皆為 1/2。只是雖說主觀，他亦能講出其依據：快艇與獨行俠，例行賽在西區的排名分居 4、5，差異不大，兩隊對戰的勝負為 1：2，差異也不大。就算要說有差異，數據上顯示兩隊各一優一劣，算是平手。再加上快艇雖有雷納德及喬治兩大球星，獨

行俠也有表現常如超人般的唐西奇，因而他覺得兩隊實力難分軒輕。可看出此君之估計雖說主觀，但也是跟據一些客觀的數據。

由上述討論知，即使擁有同樣的數據，人們對機率的判定，仍可能相差極大。且有時旁人視為主觀，自己卻認為客觀。那機率不是各說各話嗎？通常的確如此！事實上，平常諸多號稱專家，開口閉口“依據大數據…”，聽起來權威性十足，但數據躺在那裡，到底如何運用，很可能是相當主觀甚至隨意，並不見得皆具權威性。

賽前固然可大吹法螺，斬釘截鐵地做各種預測，一旦季後賽開打，預測是否準確，便要見真章了。快艇隊在例行賽的最後一場，以 5 分之差敗給西區排名倒數第 2 的雷霆隊，引起媒體奚落，說該隊“演很大”，因前一場快艇才輸給西區排名最後的火箭隊 7 分。以連敗兩場收尾，讓快艇例行賽的戰績，從原本的西區第 3 落至第 4，因而季後賽首輪會對上排名第 5 的獨行俠。當時媒體報導，若快艇“正常”出賽，將有不小機會在季後賽首輪對決拓荒者，或者衛冕軍湖人，但快艇以為獨行俠較好打(去年季後賽，快艇以 4 比 2 擊敗獨行俠)，於是快艇“故意”先後敗給西區倒數 1、2 名的火箭和雷霆兩隊。我們說過球隊可有各種策略，既然進入季後賽的目標已達成了，就開始布局下一個目標，至少能走愈遠愈好。媒體的報導，究竟是小人之心的猜測，或者揭發真相？當然就無從得知了。

媒體雖是取笑快艇，獨行俠卻顯得很不是滋味，因這表示他們被普遍認為實力遠不如快艇，只是這口氣只好暫時忍下來。季後賽首輪打了兩場後，快艇被嘲諷偷雞不成蝕把米—挑錯對手了，居然連敗兩場！ESPN 找的那群專家灰頭土臉了嗎？只能啞巴吃黃蓮，因獨行俠在系列賽的前兩場，打出令人驚嘆的表現：第一場以 47.6%之 3 分球命中率投進 17 個 3 分球；第 2 場再以 52%之 3 分球命中率投進 18 個 3 分球。這時快艇不再被看好了，有人以為他們晉級的機率，已差不多是 0 了。因如果再輸 1 場，便陷入 0：3 的落後，而 NBA 史上，在此之前，季後賽累計有 140 支陷入 0：3 落後困境的球隊，無一能逆轉成功。只是賽前不是說快艇有 17/18、1/3，或 1/2 的機率過關嗎？怎麼如今第 3 戰尚未開打，便好像被當做已輸定了，且過關的機率被認為已與 0 差不多了？要知機率值會變是機率的特性，視已知的結果，機率值可能會隨之調整，這便是條件機率。當然也有不變的情況，如知道東區某隊之戰績，對屬於西區之快艇能晉級的機率，應不至於有影響，即此二事件乃相互獨立。附帶一提，前述歷來從 0：3 逆轉成功的數據，雖是來自不同年分的不同球隊，但在沒有其他更有效的參考數據下，總是一值得參考的依據。

雖被看衰，快艇並不氣餒，由 0：2 落後連贏兩場，將戰績扳成 2：2 平手。這時快艇重獲媒體青睞，再度好評連連，這是合理的。其實 NBA 球迷一向看好快艇，2021 年 5 月中旬例行賽一落幕，依美國媒體報導，最新奪冠賠率也立

刻出爐。最被看好的球隊是籃網(例行賽在東區排第2、全聯盟第4)，其次是湖人(雖然因兩位主力球員受傷多時，他們例行賽戰績不佳，在西區排第7)，第3即為快艇，優於居第4之全聯盟例行賽戰績最佳的爵士。至於獨行俠奪冠賠率榜上之看好度僅名列第10。2：2平手後，第5戰唐西奇發威，率獨行俠力克快艇，取得3：2聽牌優勢。ESPN那18位專家的預測，至此落空，因快艇已不可能在6場內淘汰獨行俠。當然快艇並非沒機會了，但必須能連勝兩場，以4：3晉級，否則便止步於第1輪。那個當初眾人不看好的獨行俠，真是讓大家滿地找眼鏡。

第6戰快艇以104：97擊敗獨行俠。退此一步即無死所，永不放棄的快艇主將雷納德大爆發，拿下45分。他全場投25(包含3分球)中18，命中率高達驚人的72%，另加上4罰全進。NBA季後賽史上，他是第4位在球隊面臨淘汰時，拿下至少45分且命中率在7成以上的球員。罕見的紀錄，讓快艇免除以2：4飲恨，而將系列賽拉抬至3：3平手。如果雷納德在第6戰少進4球，則命中率仍有5成6，還是不錯，但便將少得至少8分，如此快艇便輸了。只是球賽當然無法有“如果”，那4次投籃不進後，怎知不會發生什麼對快艇有利的事？第7戰將一決生死。ESPN可能不必大費周章，再找任何專家來投票了。條件機率，經由這6場比賽下來，不論專家或非專家，大約都會覺得在這系列賽中，兩隊勢均力敵，因此應不必再分析了，藉投擲銅板預測那隊會晉級便可，反正兩隊晉級的機率，離1/2都不會太遠。但說不定也

有人以為，球隊在主場向來較具優勢，但過去 6 戰，主場球隊全都輸球(創下 NBA 季後賽紀錄)，顯示兩隊都不擅長打主場，第 7 戰回到快艇主場，數據會說話，不是該下注獨行俠了嗎？

第 7 戰獨行俠的唐西奇大發神威，獲 46 分，14 助攻及 7 籃板的驚人表現，另外，個人及球隊破了好幾項已高懸多年的 NBA 紀錄。快艇則轟進了 20 顆 3 分球，之前季後賽搶 7 大戰的 3 分球命中數之紀錄，是由勇士和火箭隊，分別在 2016 及 2020 年所創的 17 顆(不過季後賽單場 3 分球的紀錄是 25 顆，由騎士隊在 2016 年所創; 而例行賽單場 3 分球的紀錄是 29 顆，由公鹿隊在 2020 年所創)。兩隊連連打破紀錄，卻只能有 1 家歡喜。此系列賽首度由主場球隊獲勝，幸運的快艇以 126：111 擊退獨行俠，挺進西區第 2 輪。ESPN 的專家可鬆了一口氣，4 勝 3 敗，比他們預測的 4 勝 2 敗有點誤差，但誰贏倒是講對了。

要知在壓力下，潛力是可能被激發出來的，只是雖人們皆知球賽完全不像持續投擲銅板，並不屬於重複觀測，卻仍常採用相對頻率的概念，以預測比賽結果。只能說機率原理在實際應用時，常不得不忍受誤差。雖誤差可能很大，但總比藉卜卦來預測好。無論如何，職業球賽裡，有各式各樣的數據，不看球賽轉播沒關係，但藉由分析比賽結果，將更能了解機率的涵義。

2 機率的涵義

我們現在知道了，機率有一些不同的涵義，底下將介紹幾個較常見的。首先機率是針對隨機現象而言。A 君參加一梭哈(Show Hand)遊戲，一副 52 張的撲克牌，每人發 5 張，並觀測 A 君會拿到什麼樣的 5 張牌？這種事先不能預知結果之觀測(或說試驗)，便稱為一隨機現象。甚至，為了包含更廣，即使事先能確定結果之觀測(如投擲一兩面皆為正之銅板，並觀測會出現什麼面)，我們也視為一隨機現象，可稱為退化的隨機現象。對一隨機現象，所有可能出現的結果之集合，稱為樣本空間，常以希臘字母 Ω 表之。Ω 中的每一元素，稱為一樣本，一樣本即表一觀測之可能的結果。而若干個(0 個亦可)樣本的集合(即 Ω 之一子集合，空集合$\emptyset$亦可)，便稱為一事件。對投擲一銅板，樣本空間不大，只包含正面及反面兩個元素，以{正面，反面}，或就以{正，反}表之。至於對一梭哈遊戲，樣本空間就很大了，包含 C(52,5)=2,598,960 個，即共有兩百多萬個元素。

先看“古典的模式”，此乃以“相同的可能性”來解釋機率。古典的模式約產生於 17 世紀機率的萌芽期，這是“古典”一詞的由來。當時有些人會絞盡腦汁，求在梭哈遊戲(或其他賭戲)裡，各種事件發生的機率。如拿到葫蘆(Full house，5 張牌裡有 3 條及 1 對)之機率為何？同花(Flush，5 張牌同一花色)跟 4 條(Four of a Kind，5 張牌裡有 4 張同樣的點數)那

一個大(機率愈小的事件便愈大)？不必多想便知，那時計算機率者，心中只會有相同的可能性。即想當然耳地以為，有辦法先將牌洗得無比均勻，使兩百多萬個可能的組合，出現之機率皆相同。如此利用排列組合，便求出拿到葫蘆之機率為 3,744/2,598,960，即約為 0.00144。又 4 條比同花大，兩者發生之機率分別為 624/2,598,960 及 5,108/2,598,960。如果牌洗不均勻怎麼辦？那時是無法考慮這種問題的。其他諸如銅板及骰子等，在機率發展的初期，往往不必說明，就逕自假設是公正。在古典的模式裡，若樣本空間裡有 n 個元素，即一觀測總共有 n 個可能的結果，便將每一結果出現的機率皆當做 $1/n$。而對一包含 k 個元素之事件 A，A 發生之機率，以 $P(A)$表之，便為 k/n。

古典的模式並非通行無阻。例如，在 NBA 裡，數據就在那裡，怎還會假設 30 支球隊實力相當，因而每支球隊奪得年度總冠軍之機率皆為 1/30？這種例子比比皆是，如銅板不必然就是公正，而自有良品及劣品之某批產品中任取一件，若良劣所佔比例不同，則會取中劣品之機率，豈能就當做 1/2？既然古典的模式無法適用於所有隨機現象，其他機率的涵義便因應而生。

對一可重複觀測的隨機現象，如果對其中某事件 B 有興趣，則於觀測 n 次後，若 B 出現 k 次，我們便以 B 出現的相對頻率 k/n，當做 B 之機率。此稱為以頻率來解釋機率，算是一種客觀的解釋。說客觀乃因就是重複觀測，讓數據說話，與誰來觀測無關。此種機率的觀點，稱為頻率的觀點，

或客觀的觀點。直觀上，當觀測次數 n 愈大，相對頻率 k/n，就愈接近 B 之機率 $P(B)$，只是實際上卻未必真會這樣。不同的人(或同一人在不同的時候)同樣觀測 n 次，B 出現的次數，很難每回都相同。換句話說，B 出現次數之相對頻率是波動的，從 0(對應 k=0)至 1(對應 k=n)都有可能。即使觀測次數 n 再大，k/n 也不一定會很接近 $P(B)$。我們該修正一下前述直觀：只要 n 夠大，要多相信(即機率可任意接近 1)相對頻率 k/n 會多接近 $P(B)$(即$|k/n-P(B)|$可小於任一正數)，都辦得到。這樣講，或許有人覺得很抽象，但這涉及極限的概念，一時並不易解釋清楚，我們就此打住。

有客觀顯然便有主觀，隨機現象不乏無法重複觀測的。如中美大戰發生之機率為何？男孩C追上心儀女孩D之機率為何？這類事件之機率，恐怕皆無法經由重複觀測得知，往往只能主觀地認定，因而稱為主觀的解釋。此種機率的觀點，可稱為主觀的觀點，而所得之機率值便稱為主觀機率。對同一事件，不同人之主觀機率可能差異很大。如 C 信心滿滿地認為，追上 D 之機率為 0.8，此乃因 C 注意到，D 看到他都會微笑，顯然對他有好感。只是了解 D 之 C 的好友 E，卻認為 C 追上 D 之機率最多只有 0.001，因 D 向來對人友善，看到人微笑，根本不足為奇，加上 D 已有相識多年的男友，目前一點都看不出 D 有移情別戀之跡象。另外，即使對能重複觀測的隨機現象，主觀機率仍不時能派上用途。如 F 就是堅定相信銅板是政府發行的，不可能偏頗，故投擲出現正面的機率，必為 0.5。雖主觀機率也可能基於客觀的數據，只

是對相同的數據，不同的人可能得到不同的機率。如對中美大戰發生之機率，m 個專家之判斷便可能有 m 個。

雖說“古典”，至今仍經常出現在對機率的解釋中。例如，某國家考試，錄取並非依抽籤，每個人考上的機會自然有大有小，但由於報名者眾，錄取名額寡，在沒更好的辦法下，不少人會就簡單地由錄取率，來估計任一報名者之錄取機率。又，我們介紹的幾種機率之涵義，並非互不相容，有時會混著運用。如 G 校參加某項排球賽，分組預賽後，取 6 隊進入複賽，再取兩隊打決賽。G 校了解自己實力不錯，初賽時主觀覺得己校能晉級之機率為 0.8。6 校進入複賽後，採單循環賽。G 校覺得各校有不同策略上之運用，勝負難料，認為自己能進入決賽之機率為 1/6，古典機率出現了。G 校順利進入決賽後，查出過去一年，曾與對手交戰 8 次，5 勝 3 負，遂預估能獲冠軍之機率為 5/8，這裡便用到頻率的觀點。

有人會說，在實際應用時，上述 3 種機率的涵義，都不能顯現精準性。例如，銅板怎確定毫無偏差？撲克牌又如何得知已洗均勻了？因而運用古典機率時，不是太草率嗎？甚至即使同一銅板，每次投擲的力道總會有差異，出現正面的機會不必然皆一樣，因而視為每次投擲一不同的銅板，說不定還較恰當些。更不要說以 18 個不同的人，其中 17 位支持某事件發生，就以 17/18 當做該事件發生的機率，意見來自不同的人，這豈能算是頻率的觀點，豈能宣稱是客觀的機率？至於主觀機率，既號稱主觀，固然可不追究機率的產生是否合理，但心裡想的機率，較可能是諸如 0.2、0.7 等，較

簡單的數值，而不會有像是 0.00144 這類數值，這不是太粗枝大葉嗎？

鼎鼎大名的法國數學家及天文學家拉普拉斯(Pierre Simon，Marquis de Laplace，1749−1827)，對機率的重要性極為推崇。他曾說，“這門源自考慮賭博中的機運之科學，必將成為人類知識中最重要的一部分，生活中最重要的問題中的大部分，都將只是機率的問題。”只是機率就算再怎麼重要，前述質疑仍皆有道理。即高高在上的機率，一旦落入凡間，的確常會顯得左支右絀，揮灑不自如。我們說過了，很多事件的機率，只有天曉得。但對某一有興趣的事件，若有人只是想粗略地知道其發生的機會有多大，那只好藉助上述幾種機率的簡單詮釋法。既然以簡易的方式來探索一未知的量，便不得不忍受難以避免的誤差。

事實上，3 種機率的涵義，還有本質上的不足，那是更不容忽視的。例如，大家可能已看出來，在古典的模式裡，可能性之多寡(即樣本空間中的元素個數)，隱含假設須為有限。即不能求諸如從自然數中任取一數，會取中 3 之機率。直觀上，此機率應為 0，因自然數有無限多個。我們甚至知道，任一自然數，會被取中之機率皆為 0。只是若真的去取，必會取中某一自然數，但不是才說任一自數被取中的機率皆為 0，但機率是 0 的事件怎會發生？再仔細想，任取一自然數又是什麼意思？如何取？這些問題，在古典機率的時代，是無法解釋的。另外，在古典機率裡，所用到的工具，大抵就只有排列組合。不能討論極限，而所會用到的工具排列組

合，在數學裡被認為頂多是計算有時複雜些，但一點都不深奧(僅屬於高中數學的範疇)。基於這些原因，一般數學家，怎會認為機率裡有什麼了不起的學問？

ESPN 找來的那幾位專家，究竟功力如何？不知道！機率從來不是僅由幾次觀測的結果，就能論斷準確性。必須觀測多次，待數據夠多後，才能來檢驗。若對某事件發生的機率懷疑將如何？例如，有一原本被視為公正的銅板，但由投擲的經驗(頻率的觀點)，懷疑該銅板出現正面的機率不為 0.5；或懷疑(主觀的觀點)某廠牌手機電池，怎可能如其宣稱可待機長達 10 天。這些是常有的情況。再如某君施打了某廠牌之新冠肺炎疫苗，後來卻發現確診。有人解釋，此廠牌之疫苗，其療效之免疫力雖已達 95%，但仍有 5%會失效，該君就是屬於那倒楣的 5%。真是這樣嗎？統計裡早已發展出一套假設與檢定的程序，可對懷疑的某假設做檢定。只是隨機世界裡常是真假難辨，銅板是否公正？只有天曉得！一切都是假設，就看接受那一個？

I 女無法確定正交往中的 H 君，是否為她適合之對象？她遂先假設 H 君適合(秉持無罪推定的精神)，然後開始觀測 H 君的表現，看他能否通過她所設定的條件。不難理解，一假設能否被接受，與條件有關，條件愈寬便愈易接受。眾所皆知，I 女一旦接受 H 君，並不保證就找到真命天子。統計上，對一檢定，接受它，不表該假設必為真。只表在給定的條件下，能接受該假設。此正如擇偶一般。

3 公理化的機率

什麼是孝？在“論語”“為政篇”裡，孔子針對 4 個不同學生之提問，分別給出不同的回答。1 個不夠，要將 4 個回答一起看，才能理解孔子心目中對孝的看法。很多概念可能都是這樣，在不同的情境下，有不同的解釋。至於我們之前介紹的 3 種機率之涵義，通常已能包含生活中大部分會遇到的隨機現象，其中所涉及的機率之意義，因而本來這 3 種對機率的解釋已足夠了。但數學裡追求的，或者說數學家探討的主要興趣，往往不是為了生活中之需要。就算問題始於實際生活中之遭遇，數學家向來習於一般化，進而追求抽象化，所得結果力求包含更多的情況。像是對於解一個特別的 1 元 2 次方程式，根本不以為值得一提，至少要給出一般式的解。積分也一樣，求出圓面積及圓錐體積等，其實已不容易了，卻仍難以滿足數學家的好奇心。不能只對一些特殊的圖形求其面積或體積，要就解決一般函數如何積分，而非求出一個又一個的特別函數之積分。

我們知道，人們最早理解的數是自然數，進而掌握整數及有理數，再來發現無理數的存在，因而了解實數系統。數學家融會貫通後，便引進以公理化的方式，來定義實數系統。也就是先給一集合，有人好奇此集合裡的元素究竟是什麼？無關緊要，連是否為數字都不用在意。然後對集合中的元素定義二運算，且給出該二運算必須遵循的 10 條公理

(axiom，或說規則)。既然稱為實數，該二運算是否一為加法，一為乘法？很多人再度感到好奇，而數學家也再次指出，這一點都不重要，看此二運算滿足什麼條件才重要。要知“名”僅是表象，正如玫瑰之所以受人喜愛，乃因其優雅的外觀，及不同的花色能喚起人們不同的情感，而不是因為它叫玫瑰。但於細心觀察後，將可發現，我們原有的實數系統，完全滿足該公理化的實數系統。至於那兩個運算，可一個對應加法，一個對應乘法。也就是人們熟知的實數系統，乃為公理化的實數系統之一特例。

現在來看，如何以公理化的方式，來引進機率？我們先想，已知的 3 種機率的涵義，其中有那些要件？首先要有觀測的結果。我們遂給一集合 Ω，稱做樣本空間。Ω 便對應某一觀測之所有可能結果之集合。這集合要滿足什麼條件嗎？有，但實在不多，只要不是空集合即可。因一個什麼結果也沒有的觀測，豈值得討論？那可以只包含 1 個元素嗎？是可以，只是若只有 1 個可能的結果，不就是在觀測一退化的隨機現象嗎？通常人們對此興致不高。又，一定真要觀測什麼？那倒是不必，所謂觀測可以是虛擬的。就像有人講故事時，常以“從前有個美麗的公主”為開頭，既然是故事，便不必去追問“從前是什麼時候？”或“真有這個公主嗎？”。因而完全可以不必理會是否真有什麼觀測。

要談機率須有另一要件，那便是事件。所謂事件，乃我們有興趣，想給它一機率值之 Ω 的某一子集合。將所有事件聚在一起，構成一集合。此集合要滿足什麼條件嗎？有，就

是我們要有足夠的包容性。有點像在擬婚宴賓客名單時，賓客可以很少，少到除男女雙方的父母外都不請，否則每增列 1 人，便可能會隨之增加幾個人。依所訂的規則，充實此事件之集合(每一事件皆為 Ω 之一子集合)$\boldsymbol{F}$。如果不願多花腦筋，將 Ω 之所有子集合全找出來，所構成之事件的集合，也是可以。例如，假設 Ω 中有 n 個元素，則 Ω 共有 2^n 個子集合，隨著 n 之成長，2^n 成長快速(n=10 時，2^n 已是 1,024)。但有時我們可能覺得，不必有這麼多事件，太累贅了。也就是說，表面上看來，我們是在對事件的集合設關卡，其實不然，乃是讓事件的集合更一般，可大可小。正如對於婚宴賓客名單，要求所有朋友全都要請，跟要求若請某群死黨中的 1 位，則這群人中的其他死黨也都得請，當然是後者較有彈性。例如，NBA 有 30 支球隊，A 君只對公鹿隊有興趣，想知道公鹿奪冠的機率，所以{公鹿奪冠}須為一事件。但豈能當鴕鳥，不想知道公鹿不奪冠之機率？所以{公鹿不奪冠}也須為一事件。而若不但想知道公鹿奪冠的機率，也想知道太陽隊奪冠的機率，則要求也得知道公鹿或太陽奪冠的機率，即{公鹿奪冠，太陽奪冠}也須為一事件。當有興趣的事件愈來愈多，事件的集合便愈來愈大了。

最後須有一機率函數 P。所謂機率函數，乃對 $\boldsymbol{F}$ 中的每一事件 A，給一 $P(A)$，當做事件 A 之機率。機率函數，既然以機率之名，當然要符合人們向來對機率的認知，即須滿足一些基本的條件，如 $P(A)$須介於 0，1 間等，否則便無法稱為機率了。樣本空間、事件的集合，及機率函數，三者便構

成機率空間$(\Omega,\boldsymbol{F},P)$。至此大家可看到，雖說抽象化，機率空間其實並不真那麼抽象，至少比有些經文具體多了。如金庸(1924−2018)武俠小說裡，宣稱練成後威力無窮的“九陰真經”，便處處難以體會。像是那句“天下之至柔，馳騁天下之至堅”，簡潔有力，令人不禁躍躍欲試。但到底如何馳騁呢？就沒說了。

已經運用幾百年的機率，為何須公理化，引進機率空間？首先公理化就是孔子講的“一以貫之”。不論採何種方式解釋機率，都可在機率空間裡，找到他以為的機率。另外，這樣抽象化後，再無銅板、骰子，或撲克牌等，便能討論較一般的問題，且樣本數可有無限多個。更重要的是，有夠多的理論可探討了。其實，與數學的其他領域相比，機率論的發展晚很多。我們說過，有很長一段時間，大部分的數學家，不過視機率為很低階的數學，並不太放眼裡。就是銅板、骰子或撲克牌，沒太多理論可討論。機率的公理化，乃要歸功於俄國的科莫果洛夫(Andrey Nikolaevich Kolmogorov，1903−1987)。教科書裡具開創性的數學家，大部分都是幾個世紀前的人物。科莫果洛夫的“機率論的基礎”(Foundations of the Theory of Probability)，出版於 1933 年，不到百年前。在這本少於 100 頁的小書中，他介紹了公理化的機率，且說“機率論作為一門數學學科，可以而且應該從公理開始發展，就如同幾何與代數一樣。”才年方 30，便極具洞察力的科莫果洛夫，不卑不亢地將機率與已有兩千多年歷史的幾何與代數相提並論，進而使機率論提昇到另一層次。

公理化後，機率論便快速地發展，早已成為數學中一可與幾何與代數並駕齊驅的重要領域。只是對大部分的人來說，數不過是數字，根本沒必要理解公理化的實數系統。同樣地，大部分的人理解 3 種簡單的機率涵義，便已足足有餘了。那引進公理化的機率，究竟何用？我們說過了，對學者而言，往往會力求研究所得到的結果更一般化，成果能適用愈多的情況愈好。因此能在機率空間上探討問題，乃極大的成就。至於對一般人，當所討論有關機率的問題，淪於雞同鴨講、夾纏不清時，就要追究此處機率空間為何？正如在“莊子”“秋水篇”裡，當莊子發現，他與惠子的“魚樂之辯”，已是各說各話時，他說“請循其本。…。”機率裡的請循其本，就是找出機率空間。

由於樣本空間 Ω 中的元素不一定是實數，而人們較習慣處理實數，因而引進隨機變數(random variable)，即一從 Ω 對應到實數集合 R 的函數，不少人學機率時，便是從這裡開始迷失了。明明是一函數，卻稱變數，而且數學裡，函數通常以小寫英文字母 f、g、h 等表之，但對隨機變數，卻通常以大寫英文字母 X、Y、Z 等表之。又對一函數 f，運算時，常寫成 $f(x)$，其中 x 表其變數，但對一隨機變數 X，卻常就單單寫 X，較少附帶其變數。

另外，即使樣本空間 Ω 中的元素本就是實數，也常引進隨機變數。如令 X 表某一觀測的結果，而若觀測 n 次，則可令 X_1，X_2，X_3，…，X_n，分別表 n 次觀測所得的結果。對一隨機變數 X，$P(X\leq x)$，表事件 $\{X\leq x\}$ 之機率，而函數

$F(x)=P(X\leq x)$，$x\in R$，便稱為隨機變數 X 之分佈函數。實務上，常遇到的隨機變數 X，取值有兩大類。第一類是離散型，如 X 取值在有限集合$\{0,1,\cdots,100\}$，或無限集合$\{0,1,\cdots\}$等，這時函數 $p(x)=P(X=x)$，其中 x 為 X 可能取的值，便稱為 X 之機率密度函數。第二類是連續型，如 X 取值在區間$[0,1]$、$(0,\infty)$，或 R 等，此時若有一函數 $p(x)$，使得 $P(X\leq x)=p(x)$在區間$(-\infty,x]$之積分，則 $p(x)$，$x\in R$，亦是 X 之機率密度函數。給了機率密度函數後，分佈函數便決定了。不同的隨機變數，可有相同的分佈函數，或就簡單地說有相同的分佈。但有相同的分佈之隨機變數，卻可能相差十萬八千里。有一些常見的分佈函數，待有機會再介紹。正如人的同名同性，兩個毫不相干之隨機變數，亦可能會有相同的分佈函數。

對一離散型的隨機變數 X，假設所取可能的值為 x_1，x_2，$\cdots$，x_n，則 $x_1f(x_1)+x_2f(x_2)+\cdots+x_nf(x_n)$，便稱 X 之期望值，常寫成 $E(X)$，且常以希臘字母 μ 來表示。若 X 可取無限多個可能的值，此時前述和便無止盡地延伸，而期望值就不一定存在了。又$(x_1-\mu)^2f(x_1)+(x_2-\mu)^2f(x_2)+\cdots+(x_n-\mu)^2f(x_n)$稱為 X 之變異數，寫成 $\mathrm{Var}(X)$，常以 σ^2 表之，其中 $\sigma>0$，為一希臘字母。再度，若 X 可取無限多個可能的值，此時前述和便無止盡地延伸，那變異數就不一定存在了。至於連續型的隨機變數 X，其期望值及變異數，分別為 $xf(x)$，及$(x-\mu)^2f(x)$，在取值範圍內之積分，也都不一定存在。又，變異數之正平方根 σ，即為標準差。期望值及標準差，是一分佈之兩個相當重要的值。

4 機率實作

球隊比賽，有時以投擲銅板來決定攻防先後，此因人們多半視銅板為公正。生活上，銅板除當錢幣外，也偶被用來當做公正的象徵。若接受某一銅板為公正，那便會相信，只要持續投擲該銅板，則正面數出現的相對頻率，將很可能愈來愈接近 0.5。這便是著名的大數法則(law of large numbers)。實踐檢驗真理，真有人動手執行過實驗嗎？

坐牢時能做什麼？大部分的人是沒辦法在獄中完成“正氣歌”之類的傳世作品，或其他什麼偉大的成就，只能長吁短嘆吧！但其實仍有些事可以做。克里奇(John Edmund Kerrich，1903−1985)出生於英國，但在南非成長，後來回到英國唸大學。自 1929 年起，他開始擔任數學講師。他稱不上什麼了不起的數學家，使他出名的，是因他曾執行一系列的機率實驗。1940 年 4 月初，他到丹麥首都哥本哈根拜訪親戚，在預計返回英國的兩天前，4 月 9 日，德國納粹入侵丹麥。戰事僅持續 6 小時，丹麥政府便宣告投降，所以此戰役有時被稱為六小時戰爭(Six Hour War)，是第二次世界大戰期間，為時最短的戰役之一。南部與德國接壤的丹麥，對德國的戰略價值並不高，德國佔領的主要目的，是用來作為入侵與丹麥一海之隔，戰略價值極高的挪威之集結地，並確保部隊補給線之安全。至於挪威戰役(Norwegian Campaign)，則從 1940 年 4 月 9 日起至 6 月 10 日止，為期超過兩個月，雖最

後仍淪陷，挪威可說奮勇抵抗。

在被囚禁期間，閒閒沒事幹，克里奇遂以銅板及乒乓球，來展示幾個簡單的機率法則之有效性。克里奇將每次投擲結果均記載下來，第二次世界大戰後，他出版“機率論實驗導論”(An Experimental Introduction to the Theory of Probability)，發表他的實驗紀錄。克里奇共投擲銅板 1 萬次，其中有 5,067 個正面，正面數出現的相對頻率為 0.5067。此觀測能支持該銅板為公正嗎？是有點偏差，但大家想必了解，不能要求剛好得到投擲數之半 5 千個正面，才相信銅板為公正。事實上，若每次擲銅板實驗，皆恰得到半數個正面，一般人比較可能的反應，應是懷疑其中有貓膩，而不是就此心服銅板為公正。有點小偏差反而更易讓人相信銅板為公正，只是偏差可容忍到多大？

可求出當銅板為公正，投擲 1 萬次後，所得正面數之期望值為 5 千，標準差為 50。如今正面數偏離期望值 67，即偏了 1.34 個標準差。這樣的偏差算大嗎？一般會認為並不算太大，因利用中央極限定理(central limit theorem)來近似，偏差(超過或不足)至少 1.34 個標準差，其機率約為 0.18，不算難發生。若就此便推翻銅板為公正，似太嚴苛了。因如此每 5、6 個實際公正的銅板，便差不多會有 1 個將被誤判為不公正。通常大約偏差逾 1.96 個標準差，才會不接受銅板為公正，即所得正面數，大於 5,098 個，或小於 4,912 個，才會拒絕銅板為公正之假設。在銅板為公正下，此機率約為 0.05。統計實務上，不超過 0.05 的機率，通常才被認為算小的。例

如，若得到 5,100 個正面，便會拒絕銅板為公正。

上述說明顯示，若欲以投擲來估計銅板出現正面之機率，即使投擲數高達 1 萬次，其估計大致也只能到小數 1、2 位準確。這是很正常的。現在來看投擲數多達 1 百萬次會如何？假設銅板公正，則所得正面數之期望值為 50 萬，標準差為 500。所以得到的正面數，落在不超過期望值 1.96 個標準差之區間，即[499,020，500,980]，其機率約為 0.95。此時對出現正面機率之估計值，將有約 0.95 的機率，介於 0.49902，至 0.50098 間，大約到小數點後第 3 位準確。此顯示估計要夠準，所花的代價是投擲數要相當大。即使這樣，仍有約 0.05 的機率，估計值不落在前述區間。由此可看出統計與數學之別：在數學裡追求精準，機率值差 0.05 算是相當大。但統計裡，若誤差的機率為 0.05，並不會覺得過大。尤其若數據本身就有難以掌握的誤差，如做民調時，樣本品質往往不會太高，此時追求過小的估計誤差，或誤判機率值，反會擔心造成他處更大的誤差。所謂明足以察秋毫之末，而不見輿薪，如此將得不償失。

歷史上，曾實際去多次投擲銅板的人，可能並非太多，因那畢竟是一件相當單調的事，且數學家對於實作，通常興趣不太大。近年來則有一受人關注的“投針實驗”。

數學裡有幾個既特別又重要的常數，其中圓周率 π，應為人們最熟知的一個。圓周率乃圓周長與其直徑之比，而半徑是 r 的圓，周長則為 $2\pi r$，面積為 πr^2，這是人們從小學起

便知道的。又，以希臘字母 π 來代表圓周率，乃英國威爾斯(Wales)的數學家瓊斯(William Jones，1675−1749)所提議的。不過歷史上，至少自阿基米德(Archimedes，西元前 287−212 年)的時代起，便持續有人探討 π 了。而早在距今 1 千 5 百多年前，我國的祖沖之(429−500)，這位南北朝時第一個朝代宋的著名數學家，約於 480 年，利用我國另一位重要要數學家，三國時代魏國劉徽(約 225−295)的割圓術，從圓的內接正 6 邊形開始，持續分割，至正 24576($=6\times2^{12}$)多邊形，求出

$$3.1415926 < \pi < 3.1415927,$$

精確到小數點後第 7 位。π 的近似值常簡單地取為 3.14，或到小數點後第 5 位的 3.14159，或更精準的 3.14159265358979 等。π 乃一無理數，甚至是超越數(transcendental number)，即不為一整係數多項方程式之根。若以分數表示，則 π 有 22/7、333/106、355/113、52,163/16,604、103,993/33,102、及 245,850,922/78,256,779 等之表示，依序分母愈大愈精準。

現在便來看，什麼是布豐投針問題(Buffon's needle problem)。布豐(Georges-Louis Leclerc，Comte de Buffon，1707−1788)是 18 世紀一位涉獵相當廣泛的法國科學家。他曾提出一投針問題，我們將問題稍修飾如下：假設一平面上有無限條間距為 l 之平行線，隨機地投擲一長度為 a 之針至此平面上，其中 $a<l$。求針會與其中某一條平行線相交之機率。底下略說明解法。

將平行線的某一方向稱為左，另一方向稱為右，而與平

行線垂直的方向，一稱為上，一稱為下。令 X 表針落在平面後，與右方向之夾角，夾角落在區間$[0，\pi)$，又令 Y 表針下方端點，與最接近的上方那條平行線之距離，此距離落在區間$(0，l]$。而所謂隨機地投擲一針至平面上，我們解讀成 X 與 Y 獨立，且分別在區間$[0，\pi)$及$(0，l]$均勻分佈。即 X，Y 之聯合機率密度函數

$$f(x,y)=1/(\pi l)，x\in[0,\pi)，y\in(0,l]。$$

則

$$P(\text{針與某一平行線相交之事件})=P(Y\leq a\sin X)。$$

上式右側機率，即 $f(x,y)$在 x-y 座標平面上長方形區域$[0,\pi)\times(0,l]$之重積分。經簡單的計算，得

$$P(\text{針與某一平行線相交之機率})=2a/(\pi l)。$$

至於上式左側機率，可經由投擲針 n 次，若其中有 k_n 次針與平行線相交，則以相對頻率 k_n/n 來估計。於是當 n 夠大，k_n/n 便可用來近似 $2a/(\pi l)$，即 π 可以 $2an/(lk_n)$來近似。這便是以投針來估計 π。至於這樣的估計能有多準，當然與 n 之大小有關。而且，如前述對以投擲來估計銅板出現正面機率之說明，若欲對 π 的估計夠精準，n 通常須相當大才行。

上述近似過程，乍看之下有些神奇，因圓周率為一常數，如今卻能以一套隨機的程序來估計它，且這豈不就是數學與統計的結合嗎？令人振奮！其實一特定銅板出現正面

的機率，也是一定值，只是未知而已，而人們也是以投擲銅板這種隨機程序來估計它。本質上此二估計，乃在執行類似的工作。

連政治人物，都感受到 π 的重要。2009 年 3 月 12 日，美國眾議院通過，將每年的 3 月 14 日訂為“圓周率日”(Pi Day)。聯合國科教文組織則於 2019 年通過，每年 3 月 14 日為國際數學日(The International Day of Mathematics)。當然這天各國應都不會放假。如何慶祝國際數學日？自 2020 年起，中華民國數學會於 3 月 14 日，舉辦“萬人布豐投針實驗”。依該學會之報導，2020 年全台共有兩百多個班級參與，投擲數最多的是 10,040 次，測得圓周率為 3.4451；投擲數最少的是 224 次，測得圓周率為 3.0046。全部回收的數據(排除 14 份異常值)中，總投擲 229,105 次，其中相交 138,955 次，依布豐公式，得到圓周率之估值為 3.140516⋯，絕對誤差約千分之一，相對誤差約萬分之 3。至於 2021 年，除了事先邀請全台 152 所學校，超過萬名師生參與布豐投針實驗，也廣邀民眾，於 3 月 14 日當天，在國立臺灣科學教育館(簡稱科教館)，一起參與現場的投針實驗。152 校師生，事先之投針實驗合計後，估算出的圓周率為 3.1372；科教館現場民眾參與的實驗，估算出圓周率則為 3.1418。

2019 年 3 月 25 日聯合報的好讀周報有則新聞，開頭便說“Google 本月 14 日宣布，日籍女工程師艾瑪岩尾(Emma Haruka Iwao)耗時 4 個月，以雲端服務算出圓周率至小數點後 31.4 兆位數，刷新此前 22.4 兆位數的紀錄，以此紀念 π

對科學發展的重要性。”31.4 兆位，就是 31.4 萬億位。注意這是“位數”，而 1 兆“僅”有 13 位。若以 A4 紙列印，假設 1 頁能印 4 千位，需要 78.5 億頁，才能印出這麼長的一串 π 值。若 1 本書有 1 千頁，將需要 785 萬本書，才能印出來。在 2019 年的圓周率日，Google 公司宣佈打破紀錄，求出圓周率至小數點後 31.4 兆位，以為慶祝。而我們如何慶祝？大費周章，卻只能求出圓周率至小數點後 1 或 2 位。參與投針實驗的萬名師生，因此對機率的認識有增加嗎？能更引起學習機率的興趣嗎？恐怕皆未必。因他們每個人所做的事，不過是擲幾根不算細的“針”(由所公佈的活動照片，看起來較像竹籤)，而以所得到的結果來估計 π，將被略具數學概念的學生識破並不實際，因得到的估計值，準確性大多不如他們自小學起即熟知的近似值 3.14，不如 3.14159 準確則是一定的。

一方面已知的常數圓周率，已有很多精準的數學公式來近似，要精準到小數點後第幾位都可以，只是所花時間長短的問題。而一未知的銅板出現正面之機率，不妨以 p 表之，既然是未知，只好以隨機投擲來估計，而即使投擲數再大，都無法保證 p 之估計值，能到小數點後第幾位準確。況且若以 n 表投擲數，數學上令 n 趨近至 ∞ 不過一句話，實際擲針時，不但 n 無法趨近至 ∞，連 n 並不算太大時，執行起來便已相當困難。另一方面，投擲銅板後，那一面朝上很容易辨識；至於擲針後，那些針與平行線相交，並非那麼容易分辨。即使平面上只有一根針，有時仍得睜大眼，才能判定該針與

平行線是否相交，何況針很多時。活動照片顯示，經大量投擲後，往往多根針交錯重疊。如此計數時不產生錯誤，根本是高難度。更不要說，可能是為了更易看清楚，一條條的平行“線”，均約有 0.5 公分寬，針的寬度加上線的寬度，將讓計算的誤差變得更大了。

要知統計是沒有辦法中的辦法，當無其他辦法時，統計便能大顯神通。若已有簡易明確的辦法求出精確值，實不必再大費周章，以誤差不小的觀測實作來估計。當然針看起來似比銅板更莫測高深；圓周率比銅板出現正面之機率也似更有學問；畫有平行線之擲針地板，比不必佈置之擲銅板地板亦似更吸引人，這些都可能是舉辦擲針活動，而不是舉辦擲銅板活動的原因。

5 條件機率(一)

我們已數度提到條件機率，簡單講，一事件在給定某條件下，會發生的機率，便稱條件機率。例如，投擲一公正的骰子，會出現點數 1 的機率為 1/6。但若有人透露，出現面之點數為紅色，則因骰子只有 1 及 4 點為紅色，故此時會出現點數 1 的機率便改為 1/2。條件機率的原理，看起來並不難。不過有時因所給的條件，是以一情境來表示，若解讀不同，所得之條件機率值，將差異不小。有時甚至不易確定，究竟那一解讀才是對的，這是條件機率常會令人對感到迷惑的原因。

先給一明確的定義。對一事件 A，給定(或說知道等)事件 B 發生後，A 發生的條件機率，以 $P(A|B)$表之，且等於 $P(A\cap B)/P(B)$，其中 $P(B)\neq 0$。舉例而言，在前述擲骰子之例裡，A 即出現點數 1，B 即出現面之點數為紅色，因 $A\cap B=\{1\}$，而 $P(A\cap B)=1/6$，$P(B)=2/6$，即得 $P(A|B)=1/2$。又在 $P(A|B)$的定義中，由於置於分母，故要求 $P(B)\neq 0$。有此要求是合理的，例如，投擲一公正的骰子，若被告知出現面之點數為綠色，求此時會出現點數 1 的機率。則只能對此問題棄之不理，因骰子並無綠色點數，P(出現面之點數為綠色)=0。給一發生機率為 0 的事件，正如來亂的，是不被允許的。

條件機率的出現，可說無所不在。自 2020 年初起，新冠肺炎的疫情，在極短時間內，便進入全球大流行，寶島台

灣也無法倖免。多位醫學或公衛學者，不斷建議宜大量篩檢。病毒豈會無中生有？自 2020 年 3 月起，更有人提議入境普篩，並實施 PCR 檢測(Reverse transcriptase polymerase chain reaction，能驗出檢體中是否含有病毒的遺傳物質)，如此才能有效防堵病毒於境外。只是此建議不但未被接納，且被某些人指責是企圖製造恐慌。從快篩、廣篩，到入境普篩等建議，就是想儘速找出感染者，以控制疫情。但由於台灣人民普遍自律性高，並善於保護自己，且政府仿上古時代，傳說中的鯀設堤治水的方式，對海外若干地區，採禁止入境措施，以將病毒阻絕於外，這對海島台灣易於執行，因而初期台灣的疫情尚非太嚴重。主事者遂信心滿滿，一直以花費太貴，且篩檢會有“偽陽性”及“偽陰性”，將浪費醫療資源的理由，極力阻擋篩檢。直到 2021 年 5 月，台灣疫情快速擴散，一時無法有效控制，而疫苗短缺的情況，又看不出何時能緩解，相關單位才不得不逐步同意篩檢。到後來連居家快篩試劑都開放了，在藥局或便利商店便可買到，採檢後若呈陽性，就要盡速到醫院再做 PCR 檢測。而在各界呼籲了一年多後，2021 年 6 月底，相關單位才同意入境普篩。

偽陽及偽陰到底是什麼？居然能讓有些人如臨大敵般地藉此阻擋篩檢。

在電影或小說裡，偶會有女子拿著驗孕棒進入浴室，藉尿液來檢驗自己是否已懷孕之情境。檢驗有兩個結果，不妨分別以陽性及陰性稱之，前者表示已懷孕，後者表示未懷孕。若檢驗呈陽性反應，就會以為真懷孕了嗎？恐怕未必。

一般會至醫院做進一步檢查以確認，檢查後若證實並未懷孕，則此陽性即偽陽性。同理，若檢驗呈陰性反應，並不代表就一定沒懷孕，驗孕時間太早、尿液檢體太稀，甚至操作錯誤，都可能影響檢測結果。若其實是懷孕，則此陰性便是偽陰性。擔心偽陰性者，可以多驗幾次以確定，這是何以有人驗孕棒一次並不只買一支。在家便能做的簡易檢驗，豈能要求百分之百可靠？但使用驗孕棒這類器材，畢竟是一相當方便的驗孕方式，視呈陽或呈陰，再進一步研判。所以雖可能有偽陽或偽陰的錯誤，仍常被使用。少有人會因擔心偽陽或偽陰，而反對採用驗孕棒。

廣義而言，不但大多數的檢驗(包括酒測)、各種考試測驗(包括智力測驗及心理測驗等)，甚至生活裡的各種評比，都有相當於偽陽或偽陰之誤。即使法官也非個個是包青天，能我心如秤，總會有誤判與錯放，這是難以避免的。那不要法院了嗎？非也，就設置補救措施。如今怎麼醫學裡一項病毒篩檢的偽陽偽陰，會被某些人避如蛇蠍？

假設某傳染病之盛行率為 10%，某區域住有 1 萬人，且全參與該病之篩檢，又設篩檢的準確率為 98%。不利用條件機率的公式了，底下我們以直觀且簡易的方法，來看偽陽及偽陰的問題。依假設，該區域“平均”有 9 千人未感染，及 1 千人已感染。此處必須說明一下。要知盛行率 10%，乃表每人會感染的機率為 0.1；而篩檢準確率 98%，乃表篩檢後，每位已感染者，有 0.98 的機率呈陽性反應，且有 0.02 的機率呈陰性反應，另外，每位未感染者，有 0.98 的機率呈陰性

反應，且有 0.02 的機率呈陽性反應。對於隨機現象，談的是機率，因此我們才用“平均有 9 千人未感染，及 1 千人已感染”。不過為了簡便，底下皆略去“平均”一詞。經篩檢後，已感染的 1 千人中，有 980 人呈陽性，20 人呈陰性；未感染的 9 千人中，有 8,820 人呈陰性，180 人呈陽性。故總共呈陽性的 1,160(=980+180)人中，有 980 人真已感染，至於另 180 人便是偽陽者，180/1,160，約 15.52%；而呈陰性的 8,840(=20+8,820)人中，有 8,820 人真未感染，至於另 20 人便是偽陰者，20/8,840，約佔呈陰性者的 0.23%。而呈陽性者，僅佔總人數的 11.6%，若都去做進一步較精細的檢驗，則其中將有 15.52%的檢驗會是虛工(這應就是反對篩檢者所認為的浪費醫療資源)。雖有虛工，但若要全部 1 萬人去做較精細的檢驗，可能相當不容易，而要已呈陽性者去做進一步的檢驗，他們的意願將會高多了，且由 1 萬人降至 1,160 人，醫院的負荷也小很多。至於偽陰者，不論人數或所佔比率其實都很低。若若擔心偽陰的問題，則有如驗孕，呈陰性者可再驗一次。任一已感染者，兩次皆驗出陰性之機率僅為 $0.004(=(1-0.98)^2)$，偽陰者很難不被校正出。偽陰者的比率很低，難道此篩檢對呈陰性者較準確嗎？倒也不是這樣。此因該病之盛行率本來就不高，僅有 10%，因而任何一人屬於感染者的機率原本就較低，再經檢驗呈陰性下，會屬於未感染者的機率就更加提高(約 0.9977)。又，偽陽(180 人)及偽陰(20 人)者，共 200 人，佔總共 1 萬人中的 2%，這是對的，因篩檢的準確率是 98%。

底下將上述傳染病之盛行率降低至 1%，其他條件不變，看會有什麼改變。依假設，該區域有 9,900 人未感染，100 人已感染。篩檢後已感染者中，有 98 人呈陽性，2 人呈陰性；未感染者中，有 9,702 人呈陰性，198 人呈陽性。故總共呈陽性的 296 人中，有 98 人真被感染，至於另 198 人便是偽陽者，約佔 66.89%；而總共呈陰性的 9,704 人中，有 9,702 人真未被感染，至於另 2 人便是偽陰者，約佔 0.021%。呈陽性者，僅約佔總人數的 2.96%，若都去做較精細的檢驗，則其中將約有高達 66.89%的檢驗會是虛工。至於偽陰者，再度不但人數較少，且所佔比率更低了。又，偽陽(198 人)及偽陰(2 人)者，共 200 人，符合準確率為 98%。由本例可看出，當盛行率較低，偽陽者之比率會隨之提高，但此時篩檢仍是有效的，畢竟只需讓較少的人，去進一步做較耗工的檢驗。一般而言，對於罕見病(即盛行率較低)，受檢驗著若呈陽性，並不必過早擔憂。

現假設傳染病之盛行率 1%，而篩檢的準確率降為 95%，其他條件不變。依假設，該區域有 9,900 人未感染，100 人已感染。篩檢後已感染者有 95 人呈陽性，5 人呈陰性；未感染者中，有 9,405 人呈陰性，495 人呈陽性。故總共呈陽性的 590 人中，有 95 人真被感染，另 495 人為偽陽性，約佔 83.90%；而總共呈陰性的 9,410 人中，有 9,405 人真未被感染，另 5 人便是偽陰性，約佔 0.053%。呈陽性者，僅約佔全人口 5.90%，若都去做進一步的檢驗，則其中將有高達 83.90%的檢驗會是虛工。顯示若準確率降低，則偽陽者之比率會提

高。至於偽陰者，比率有提高些，但 0.053%仍不算高。雖此篩檢看起來對未感染者仍較準確，但此仍因該病之盛行率僅有 1%，故任何一人會未感染的機率本來就有 99%，再經檢驗呈陰性下，會未感染的機率(9405/9410≈0.9995)就更加提高了。又，偽陽(495 人)及偽陰(5 人)者，相加得 500 人，符合準確率 95%。由本例可看出，當準確率降低，雖偽陽者之比率會提高，但此時篩檢依然有相當效果。

我們再改變一下條件。假設某傳染病之盛行率 0.1%，而篩檢的準確率為 90%，其他條件不變。依假設，該區域有 9,990 人未感染，10 人已感染。篩檢後已感染者有 9 人呈陽性，1 人呈陰性；未感染者，有 8,991 人呈陰性，999 人呈陽性。故總共呈陽性的 1,008 人中，有 9 人真被感染，另 999 人為偽陽性，約佔 99.11%；而總共呈陰性的 8,992 人中，有 8,991 人真未被感染，另 1 人為偽陰性，約佔 0.011%。呈陽性者，約佔全人口 10.08%，若都去做進一步的檢驗，則其中將有高達 99.1%的檢驗會是虛工。本例顯示，當盛行率更低且準確率亦下降，將會造成更高比率的偽陽性。至於呈偽陰者，此時比率很低，僅約 0.011%，此仍是因全部的人中感染者之比率本就很少，才約 0.1%。又，偽陽(999 人)及偽陰(1 人)者，相加為 1,000 人，與準確率 90%相吻合。由本例可看出，即使準確率僅 90%，雖因而偽陽性極高，但僅須 10.08%的人再驗驗，若此傳染病極令人擔憂，則篩檢便依然值得。

讀者不妨自行改變盛行率及篩檢的準確率，將發現如同驗孕棒，在大部分合理的情況下，篩檢乃有助於找出已感染

者，且能讓人心安，並未有須大力阻擋的必要。當然若盛行率實在太低，如 0.01%(平均 1 萬人才 1 位感染)，或篩檢的準確率太低，如只有 85%(這種篩檢效果未免太差)，則可能便真不太有必要篩檢了。

6 條件機率(二)

條件機率之出現，可說無所不在。底下是一測謊之例，這當然也會有偽陽與偽陰的問題存在。某公司有一項業務機密外洩，安全主管擬藉測謊以找出洩密者，於是鎖定公司內幾位他認為較有嫌疑者來測謊。這樣做恰當嗎？

假設測謊的準確度為 9 成、有 100 人經手該項業務，且其中僅 1 人洩密。則全部 100 個經手者皆測畢後，平均將有 10.8($=1\times0.9+99\times0.1$)人顯示洩密(陽性)，但其中平均僅有 0.9 人的確洩密(也就是洩密者並不見得在那 10.8 個呈陽性者中)，另 9.9 人其實未洩密(偽陽者)。因此經測謊後，每一位可疑者，僅有 1/12(=0.9/10.8，約 0.083)的機率洩密。此顯示，主事者絕不可抱著“若沒洩密何須必擔心測謊？抗拒測謊便是心虛”之心態。另一方面，平均有 89.2 人顯示未洩密(陰性)，其中平均有 89.1 人的確未洩密，僅 0.1 人真洩密(偽陰者)。換句話說，此測謊對判定人的清白較有效。但這是合理的，因任選一人，他本來便有高機率(0.99)是清白的。由上述計算得知，經“普測”後，是有助公司縮小可疑者的範圍。但對找到真正洩密者，還有一段路要走，因此絕不能把測謊當做唯一的辦案工具。更何況，若公司並非採普測，只是主觀地挑選 1、2 個“看起來不夠忠誠的人”來測謊，並對測謊未過者高度懷疑其忠誠度，則便很可能會製造出冤屈了。因測謊準確度雖達 9 成，表面上看起來不低，但任一無辜者，

皆有 0.1 的機率測謊不通過。更何況測謊的準確度若實際沒有 9 成那麼高，或慣犯者較清白者更熟如何對抗測謊，則測謊的有效性便將更降低了。總之，測謊如同疾病之篩檢，是一種本該有用的輔助工具,但就是須善用,而非迷信其效益。

條件機率在實務上，有時會眾說紛紜，得到不同的機率值，但卻各自以為正確。其根本原因，常就是對情境的解讀不同。這原本不足為奇，就像 1950 年出品的日本電影“羅生門”，對同一情境，眾人各說各話，毫無交集，使真相難以大白。但對於涉及機率的問題，有時不易讓人領悟，也會因羅生門之故，或者根本是給的條件不夠明確，才把眾人搞糊塗，而冒然怪罪機率不可信。

先來看著名的“汽車與山羊問題”(Car-Goat Problem)。簡單描述如下：A 君被邀請參加某電視節目，現場有 3 扇門，其中 1 扇門後有汽車，另兩扇門後為山羊。A 君選擇一扇門(不妨假設是第 1 扇)，門後將是他的獎品。主持人打開另一扇門(不妨假設是第 3 扇)，見到山羊，問 A 君要不要換選？當然若要換，只有第 2 扇門能換了。“標準答案”是“該換”。不必多說明，換或不換，乃基於換選後，得到山羊的機率會不會提高。很多人覺得不用換，因原本汽車在每扇門後的機率皆為 1/3，如今知道汽車不在第 3 扇門後了，利用條件機率的思維，將得汽車在第 1 扇門及第 2 扇門後的機率，皆成為 1/2，因此換門是多此一舉。主張該換的人則認為，主持人事先必然知道汽車在那一扇門後，因而他會打開一扇後面是山羊的門(否則遊戲便進行不下去了)，

而原本汽車在第 2 或第 3 扇門後的機率，其和為 2/3，現既知汽車不在第 3 扇門後了，因此第 2 扇門後就獨自擁有全部 2/3 的機率，於是該換選擇的結論便浮出了。若實際去計算，將會印證此結論。有些主張不必換的人，堅持主持人事先並不知汽車在那 1 扇門後，他只是隨機地自第 2 及第 3 扇門中，任挑一扇打開，而剛好門後是山羊，因而換或不換，得到汽車之機率皆為 1/2。這自然也並非不可能，將門全打開後，說不定汽車果真在第 1 扇門後呢！但我們早說過，機率從來不是依少數幾次的結果，去論斷準或不準。只是主持人非為節目效果配合製作單位開門，他事先並不知汽車在那一扇門之後，只是臨時起意地打開一扇門，且剛好該門後是山羊，這種畢竟是較罕見的情況。

現在我們提出一情況，讓讀者跟“汽車與山羊問題”比較。C 君去找朋友 D 君，到後忘了門號，只知是某相鄰 3 間之一。任挑一間後，C 君正要敲門時，另兩間有間開門出來一位金髮婦女抱著一嬰兒，於是他確定 D 君不住那家。那他是否該換一家門敲，因選對的機率將由 1/3 提高至 2/3？標準答案是“不必換”。原因就留給讀者自行思考了。

再看一例。有一張在公園裡的照片，標示“史密斯家庭”(Smith family)，其中有 4 人：一中年男子、一中年女子、一站著吹泡泡且穿著裙子的女孩，及一跪在地上摟著狗穿著短褲的小孩，大半身體被遮住了，看不出其性別。問：摟著狗的小孩是女孩的機率為何？這是一有兩個小孩的家庭，男子及女子，顯然是此家庭的爸爸跟媽媽。首先若看不出兩小

孩誰大誰小，則將所給的條件，解讀為“此家至少有一女孩”，似乎是合理的。我們來看如何求解？此時問題不過是求條件機率 $P(E|F)$，其中 E 為此家兩小孩皆為女孩，F 為此家至少有一女孩。在此家有兩個小孩下，E={女女}，F={女女，女男，男女}，其中“女男”表老大是女孩老二是男孩，餘類推。又 $E \cap F=E$。在生男及生女的機率皆為 1/2，且各次生孕皆相互獨立的假設下(這是一般可接受的假設)，$P(E \cap F)=\mathrm{P}(E)=1/4$，$P(F)=3/4$，故

$$P(E|F)=P(E \cap F)/P(F)=1/3。$$

其次，若能看出吹泡泡的女孩，體型明顯大很多，應是老大，則仍有 E={女女}，但 F={女女，女男}，此時 $P(E \cap F)=P(E)=1/4$，且 $P(F)=1/2$，故 $P(E|F)=1/2$。同理，若能看出看出吹泡泡的女孩為老二，則亦有 $P(E|F)=1/2$。

若吹泡泡那小孩為男孩呢？令 G 為此家兩小孩中一為男孩一為女孩，H 為此家至少有一男孩，欲求 $P(G|H)$。因 G={男女，女男}，H={男男，男女，女男}，故

$$P(G|H)=(1/2)/(3/4)=2/3。$$

其次，若能看出吹泡泡的男孩為老大，則 G={男女}，但 H={男女，男男}，此時 $P(G|H)=1/2$。同理，若能看出看出吹泡泡的男孩為老二，則亦有 $P(G|H)=1/2$。

上述答案令有些人感到訝異，因既然生男及生女的機率皆視為 1/2，則無論如何，任一摟著狗、遮著臉、站窗簾後，

或只要看不出性別的小孩，會是女孩的機率，不都應為 1/2 嗎？結果居然與家中另一小孩的性別及排序有關！這些不同的答案提醒我們，“家裡至少有一女孩”，與“老大是女孩”，此二條件是不同的。還有，若照片未標示“史密斯家庭”，則在公園裡的那 4 人，可能毫不相干。這麼一來，因看不見容貌，故摟著狗的小孩是女孩的機率，只能當做 1/2 了。

另外，可能有人會說，既然照片中吹泡泡的女孩穿裙子，那表示史密斯家讓女孩穿裙子，則因摟著狗的小孩穿短褲，那便應是男孩，因而摟著狗的小孩是女孩的機率為 0。但這可就難說了。不過假設有人看過報導，家中若有兩個小孩，“小的”那個較會去摟抱狗，則吹泡泡的小孩便較可能是老大，這時便會傾向選 1/2 這個答案。總之，這種以一情境來表示一條件，有時的確會流於各說各話，並無標準答案。我們說過，機率並非以少數幾次來判斷準或不準，有時須經由長期觀察，才知某一情景如何解讀較可能是對的。

再看一類似的問題。有對夫妻剛搬進某社區，社區管理員 I 君，只知他們家有兩個小孩而不知性別。某日，I 君見到此家之媽媽，帶著一叫她媽媽的女孩，在社區公園玩耍。問此家兩小孩皆為女孩之機率為何？

既然看到媽媽帶著一女兒，此時求兩小孩皆為女孩之機率，不就是求留在家中那小孩為女兒之機率嗎？又因生男及生女的機率皆為 1/2，故所求之機率不應為 1/2 嗎？或者如前

述摟著狗的小孩，是女孩的機率為 1/3。事實上，這兩個答案都不見得正確。此問題比我們想像的複雜許多，關鍵在於如何解讀事件“見到此家媽媽帶著一女兒”，沒錯，此事件導致“此家至少有一女兒”，但二事件等價嗎？恐怕未必。既看到帶個女兒，則家中便不可能有 2 男；而若有 2 女，帶出門的當然是女兒，這是確定的。但當兩小孩性別不同(即有 1 男 1 女)時，在只帶 1 個小孩出門之下，媽媽究竟如何選擇帶那位？若見過統計，(1)媽媽必帶老大(或必帶老二)，此時所求機率為 1/2；(2)媽媽必帶女兒，此時所求機率為 1/3。甚至在其他不同的假設下，本問題的答案差異將極大。

我們給最後一例。有 J 及 K 二君在玩一如下賭戲：J 君投擲 2 銅板，K 君則猜兩面相同或相異，L 君則擔任裁判。不妨就假設 2 銅板皆為公正，故此為一公正賽局。賭戲開始，當 J 君擲出 2 銅板後，M 君經過，順口說“有一正面”。裁判 L 君不滿 M 君破壞賭戲之公正性。因如同上述“史密斯家庭”之例，在“此家至少有一女孩”的條件下，“此家兩小孩皆為女孩”之機率成為 1/3，如今原本兩面相同及相異的機率皆為 1/2，在 M 君說出“有一正面”(如同至少有一女孩)後，K 君便知兩面相同(兩小孩皆為女孩)之機率由 1/2 降為 1/3，因而他必會猜兩面相異，如此猜對機率將提高至 2/3。M 君不以為然，覺得自己無意也沒有影響賭戲。他說那如果他說“有一反面”，K 君不是也該猜兩面相異嗎？因猜對機率亦能提高至 2/3。但擲 2 銅板，必有正面或反面，於是只要他說話，就算 K 君沒聽清楚，只聽到“有一 X 面”，K 君

便該猜兩面相異。甚至，只要 K 君“以為”M 君有說話，在透露訊息，便都該猜兩面相異。天下居然有這種怪異的事？

事實上，M 君說“有一正面”，與說“至少有一正面”，此二事件並不等價。假設 M 君只是多嘴，卻實話實說的人。即當兩面皆為正面，他便說“有一正面”；當兩面皆為反面，他便說“有一反面”；而當有一正面及一反面時，他便各以 1/2 的機率，說“有一正面”，或“有一反面”。過程留給讀者，可求出

P(M 君說有一正面)=P(M 君說有一反面)=1/2，

P({兩面相同}∩{M 君說有一正面})
=P({兩面相異}∩{M 君說有一正面})=1/4。

故

P(兩面相同|M 君說有一正面})
=P(兩面相異|M 君說有一正面)=1/2。

因而 M 君的確並未破壞賭戲之公正性。當然亦可求出

P(兩面相同|M 君說有一反面)
=P(兩面相異|M 君說有一反面)=1/2。

也就是不論 M 君說“有一正面”，或“有一反面”，皆未提供任何有用的資訊給 K 君。

生活上處處都是條件機率，要想善用機率，弄懂條件機

率絕對是必要的。條件機率之進一步討論，可參考黃文璋(2010)，機率應用不易，數學傳播季刊，34(1)：14-28 一文。

7 隨機法則

人們常說大數據，開口閉口“利用大數據”，但有了大數據，一切問題便能迎刃而解嗎？恐怕未必。有大數據便有大數，我們先來看大數。由大數的多變，將會理解大數據的難以掌握。但有沒有能掌握時？

很多數加起來會如何？會很大嗎？不一定。對於無限等比級數，大家在中學時便知，如果公比 r 的絕對值$|r|<1$，則其和為有限。如級數 $1+2^{-1}+2^{-2}+2^{-3}+\cdots+2^{-n}+\cdots$，不論加了多少項，我們知道其和都不會超過 2，且級數和以 2 為極限。現看級數 $1+1/2+1/3+\cdots+1/n+\cdots$，一般項是 $1/n$。隨著 n 之增大，每次僅加一很小的量，由微積分知，當加到第 n 項時，$n\geq 2$，其和小於 $1+\log n$，其中 log 是自然對數的符號，即以 e 為底。所以加到 1 億項時，其和小於 $1+\log 10^8<19.43$；即使加到 1 兆項，其和小於 $1+\log 10^{12}<28.64$，仍不太大。雖成長相當緩慢，但此卻為一發散級數，其和會無止盡地增大，以∞表其和。不過形式類似的級數 $1+1/2^2+1/3^2+\cdots+1/n^2+\cdots$，一般項是 $1/n^2$，此級數和之極限可求出，為 $\pi^2/6$，約僅 1.6449，也就是除首項 1 之外，其餘無限多項的和，不超過 0.65。至於級數 $1-1+1-1+\cdots+(-1)^{n-1}+\cdots$，其和一直在 1，0 兩數間跳躍，永不會平穩下來。在微積分裡，花了不少篇幅，討論數列及級數的收斂與發散，檢驗一個數列或級數，一直持續進行下去會怎麼樣？即是否有能掌控的“終極點”(表極限存在)？或屬

於無法掌控地跳躍不止、趨近至∞，或趨近至-∞？由前述幾個例子知，在數學裡，當遇到大數，“最終”究竟會如何，有時並非那麼容易能確定。

對於隨機現象，有些特別的隨機變數之和用途不小，且大數下之行為，能顯示某種能掌握的規律。這便是所謂隨機法則，大數法則及中央極限定理，是其中極重要的兩個，我們之前已提過，底下便來加以討論。

考完試，各班成績有高到接近滿分，亦有低到接近 0 分，如何評比各班的成績？比較各班平均成績之高低，為一常見的方式。湖水多深？深淺不一，遂以平均深度來表示。其他如平均壽命、平均國民所得等，人們可說經常在求平均，並以平均當做一組量之代表值。獨立地投擲一出現正面機率為 p 之銅板，以 X_i 表第 i 次投擲的結果，$i\geq 1$，且令 $X_i=1$，或 0，分別表第 i 次投擲得到正面或反面。X_1，X_2，⋯，X_n，便是一組樣本數為 n 之隨機樣本，每一皆有伯努力分佈 $Ber(p)$。在此若 X 有 $Ber(p)$分佈，則 X 之機率密度函數為

$$P(X=k) = p^k(1-p)^{1-k}，k=0,1。$$

又可求出 $E(X)=p$。令樣本和

$$S_n = X_1+X_2+\cdots+X_n，$$

表 n 次投擲後共得的正面數，則 S_n 有二項分佈 $B(n,p)$。即

$$P(S_n=k) = C(n,k)p^k(1-p)^{n-k}，k=0,1,2,\cdots,n，$$

其中

$$C(n,k) = n!/(k!(n-k)!)，k=0,1,2,\cdots,n。$$

樣本平均 S_n/n，便常被用來估計 p。由於

$$E(S_n/n) = np/n = p，$$

所以，此估計量平均而言是準的。但平均是準的估計量很多，如 X_1 也是，因 $E(X_1)=p$。也就是若第 1 次得到正面，則估計 $p=1$，若得反面，則估計 $p=0$。直觀上，只用到 1 個樣本的估計量，不可能會好，至於 S_n/n 則會是好很多之估計量。不僅這樣，當樣本數 n 愈來愈大，利用一簡單的不等式，可得對對 $\forall\varepsilon>0$(數學裡 ε 通常表一很小的數)，機率 $P(|S_n/n-p|>\varepsilon)$，隨著 n 之不斷增大，將趨近至 0。這是什麼意思？S_n/n 取值在區間[0，1]，忽大忽小。且不論 n 再如何大，S_n/n 都"可能"(機率為正)為 0(n 次投擲皆得反面)，或 1(n 次投擲皆得正面)。但當 n 很大時，S_n/n 便不太容易(即機率很小)太偏離 p，也就是 S_n/n 大致會在 p 附近一小範圍內波動。亦即 S_n/n 偏離 p 任一正數 ε 之機率，只要 n 夠大，便可任意小。以數學式子表示，即對$\forall\varepsilon>0$，$n\to\infty$時，便有

$$P(|S_n/n-p|>\varepsilon)\to 0。$$

這便是大數法則較原始的版本。

大數法則並不侷限在只取值 0，1 的隨機變數。對一數列獨立且有共同分佈的隨機變數 X_1，X_2，…，X_n，…，若期

望值 $E(X_1)=\mu$ 存在，則對$\forall\varepsilon>0$，$n\to\infty$時，便有 $P(|S_n/n-\mu|>\varepsilon)\to 0$。這便是較一般的大數法則，適用期望值存在的任何分佈。不過這其實是弱大數法則。即有另一強大數法則，且獨立的條件亦可放寬，但我們就此打住。

2020 年美國總統選舉於 11 月 3 日舉行後，未通過連任的原任總統川普(Donald John Trump，1946–)，於 12 月 2 日，在社群媒體公布一段影片，其中呼籲翻轉大選結果。他堅稱“從統計上來說，我不可能輸”(It’s statistically impossible that I lost)。只是大家都知道，川普就是輸了。因而“不可能”究竟是什麼意思呢？

底下略說明，人們不時脫口而出之“不可能”的意思。NBA 籃球比賽，兩隊只差 2 分，領先那方投籃沒進，落後那隊搶到籃板球，時間卻只剩下 1 秒，球不得不出手了。能來個超大號 3 分球逆轉比賽嗎？距離實在太遠了，球幾乎要飛過整個籃球場，然後得進入一小籃框。觀眾莫不認為不可能，結果球居然進了，反敗為勝。對隨機現象，口語裡所謂“不可能”的事件，乃表發生機率極低。至於要多低才會被視為不可能，乃因情況而定，彈性相當大。有人問：今天會下雨嗎？答：不可能！談的機率大抵是 0.01 之類的，並不真那麼小；對統一發票，有人問：會中特別獎(要 8 碼全對，獎金 1 千萬元)嗎？答：不可能！此處談的機率為 1 億分之 1，這機率算是很小了。數學裡，圓周率 $\pi<3.1416$，這是確定的，絕無意外地成立。但對隨機現象，談的是機率。當 n 很大時，樣本平均 S_n/n 會不會沒離 $E(X_1)=\mu$ 很近？是有可能，不論 n

再大，S_n/n 都可能偏離 μ 很遠，但機率很小，且當 n 很大時，S_n/n 不離 μ 很近的機率，根本就微乎其微。微乎其微的機率何須太在意？早被歸入“不可能”的事件。因此 n 很大時，我們會視 S_n/n 差不多就是 μ 了。雖不中，亦不遠矣，而這不遠，也是指不遠的機率很大。當然，也會有人堅持認為，只要機率為正之事件，便永遠都可能發生。

其次來介紹中央極限定理。對一數列獨立且有共同分佈的隨機變數 X_1，X_2，…，X_n，…，且假設 $E(X_1)=\mu$，及 $\mathrm{Var}(X_1)=\sigma^2$ 皆存在。大數法則指出，當 n 很大時，S_n/n 接近 μ 的機率很大。也就是

$$S_n/n-\mu=(S_n-n\mu)/n，$$

此量很可能會很小(即很接近 0)。不過要注意，與 n 相比很小，絕不表 $S_n-n\mu$ 須很小。即使 n 再大，S_n 仍可能偏離 $n\mu$ 很遠。是 S_n/n 才會大致在 μ 之附近！(此正如 n 很大時，$n^{1/2}/n=1/n^{1/2}$ 的確會很小，但當 n 不斷增大時，$n^{1/2}$ 亦無止盡地增大)。有如使用顯微鏡，將草履蟲等微小物放大，以看清楚其變化，能將 $S_n/n-\mu$ 放大，或等價地說，$S_n-n\mu$ 能不要除以 n，而除以較小的量，使除之後，不至於過小也不至於過大嗎？將除數由 n 改為 $n^{1/2}$ 即可。更明確地說，標準化後的 S_n(即減去期望值再除以標準差)

$$(S_n/n-\mu)/(\sigma/n^{1/2}) = n^{1/2}(S_n/n-\mu)/\sigma，$$

當 n 愈來愈大時，其變化將愈清楚可見。而當 n 夠大時，其

分佈可以標準常態分佈 $N(0,1)$來近似。這便是中央極限定理。此處對 X 有 $N(0,1)$分佈，其機率密度函數為

$$\phi(x) = \exp(-x^2/2)/(2\pi)^{1/2}，x \in R。$$

物理上量測過程中的步驟很多，每一皆可能產生一微小的誤差。大數學家高斯(Johann Carl Friedrich Gauss，1777−1855)的誤差理論，推導出總誤差常有常態分佈。諸如身高、體重及智商等，醫學、生活及自然界裡的很多量，往往可以常態分佈當模式，常態分佈可說處處可見。這是此分佈以常態名之的主因，彷彿若其他分佈出現，便非常態。又，大數法則與中央極限定理最被人稱奇的，是其一般性。只要有一數列獨立且有共同分佈的隨機變數，什麼分佈並無關緊要，當共同的期望值存在，大數法則便適用；當共同的期望值及變異數皆存在，中央極限定理便適用。附帶一提，中央極限定理還有條件更寬鬆的版本，使其適用性更廣。不過實務上，通常上述版本就已夠了。

若想估計期望值 μ，經重覆觀測 n 次後，得到樣本 X_1，X_2，…，X_n，則可以樣本平均 S_n/n 來估計。只要 n 夠大，由大數法則知，此估計不會太離譜。但我們知道，不論 n 再大，S_n/n 並不會就一直等於 μ，至於偏離多遠？只要標準差 σ 存在，便可以中央極限定理來估計。如習慣用法，以 Φ 表 $N(0,1)$分佈的分佈函數。則 n 很大時，

(1) $P(|S_n/n-\mu|\le d)=(P(|n^{1/2}(S_n/n-\mu)/\sigma|\le n^{1/2}d/\sigma)\approx 2\Phi(n^{1/2}d/\sigma)-1$。

由上式得

(2) $P(S_n/n-d \leq \mu \leq S_n/n+d) \approx 2\Phi(n^{1/2}d/\sigma)-1$。

上式常可用來求 μ 之信賴區間(confidence interval)。底下來看一應用。

欲估計一銅板出現正面的機率 p，持續投擲 n 次，得到一數列獨立且有共同分佈的隨機變數 X_1，X_2，…，X_n，其中 $X_i=1$，或 0，分別表第 i 次得到正面或反面，$i=1$，…，n。則

$$E(X_i) = p，$$

$$\mathrm{Var}(X_i) = (p(1-p))^{1/2}。$$

故 n 很大時，由(2)式，得

(3) $P(S_n/n-d \leq p \leq S_n/n+d) = P(p \in [S_n/n-d, S_n/n+d])$

$$\approx 2\Phi(n^{1/2}d/(p(1-p))^{1/2})-1 \geq 2\Phi(2n^{1/2}d)-1，$$

此處用到，對 $0<p<1$，必有

$$(p(1-p))^{1/2} \leq 1/2。$$

若欲

$$2\Phi(2n^{1/2}d)-1 \geq 1-\alpha，$$

其中 $0<\alpha<1$，則

$$\Phi(2n^{1/2}d) \geq 1-\alpha/2。$$

現取 α=0.05，由 $N(0,1)$分佈之機率值表，得 $\Phi(1.96)\approx 0.975$，故約略有 $2n^{1/2}d \geq 1.96$。由此得

(4) $n \geq 0.98^2/d^2$，

取 d=0.03，由(4)式得 $n\geq 1{,}067.1\cdots$。由於 n 須為整數，得 $n\geq 1{,}068$。若投擲數 n=1,068，且得 S_n=551，則對 p 之估計值 S_n/n=551/1,068≈0.5159。再代入(3)式，得

$$P(0.5159-0.03\leq p\leq 0.5159+0.03)=P(0.4859\leq p\leq 0.5459)$$

$$\geq 1-0.05=0.95。$$

區間[0.4859, 0.5459]便稱為 p 之一(近似的)95%信賴區間。以一值 0.5159 估計 p(此稱為點估計)，有時會忐忑不安，覺得 p 根本“不可能”剛好等於 0.5159。但以一區間[0.4859, 0.5459] 來估計 p，看起來雖不像點估計那麼鐵口直斷，但這樣的估計卻更值得“信賴”，這是信賴區間一詞的由來。至於有多信賴呢？信心水準(或稱信賴水準、信賴度等)為 95%。

另一方面，由(4)式得 $d\geq 0.98/n^{1/2}$。有時便採 $d\approx 0.98/n^{1/2}$。或就簡單些，反正就是近似，一旦知道 n，便令

(5) $d=0.98/n^{1/2}$，

以求出估計誤差 d。因 n=1,068 時，$d\approx 0.03$，故實際操作時，如果 n>1,068，則 d<0.03；如果 n<1,068，則 d >0.03。

在上述推導中，有幾處用到不等式。如將$(p\ (1-p))^{1/2}$ 以

1/2 取代，這樣會造成誤差嗎？會。但無須擔心，因這樣將使(3)式最左側大於最右側，也就是我們得到的信賴區間，信心水準可能不只 95%，說不定還超過。而將 n 由 1,067.1…換成 1,068，又會使信心水準再稍大些。但這些其實都不必太計較，因即使 n 再怎麼大，$n^{1/2}(S_n/n-\mu)/(p(1-p))^{1/2}$ 也不會有 $N(0,1)$分佈，只是近似而已。更何況實務上，樣本數 n 通常也沒真那麼大，不過 1 千上下，因而常態近似是否有那麼準確，誰也不敢保證。於是宜保留些，將信心水準說成 95%就好了。另外，我們說有 95%的信心水準，p 會落在區間[0.4859, 0.5459]，其中 95%是指什麼？機率嗎？有人以為，在取樣前，$[S_n/n-d, S_n/n+d]$為一隨機區間，那時講 p 落在此區間之機率為 0.95 可以。但取樣後，得到 S_n 之值，前述區間成為一固定區間，此時若說 p 落在此固定區間之機率為 0.95，就沒道理了。因每次投擲會得到不盡相同的 S_n，因而得到不盡相同的信賴區間，而怎可能那些不同的固定區間，每一會包含 p 之機率，皆為 0.95？聽起來似乎有理。但別忘了主觀機率，依上述程序得到的區間，若某君覺得就是有 0.95 之機率會涵蓋 p，也沒什麼不行。甚至該君還可反問，此區間有比 0.95 更適合的涵蓋 p 之機率值嗎？

附帶一帶，不但對一固定的信賴區間，可談參數落在此區間之機率，只要未知，便可談事件發生的機率。正如懷孕 6 個月了，胎兒性別根本已定，但不知者仍可猜會生男或生女；考試當天，明明試卷已印製完成，在做最後衝刺的學生，仍可賭某題是否會考。

8 隨機法則(二)

欲估計某地區選民，對某候選人之支持度 p，遂執行一民意調查。自該地區隨機抽取 n 個合格選民做問卷，得到樣本 X_1，X_2，…，X_n，其中 X_i=1，或 0，分別表第 i 個選民支持，或不支持該候選人，i=1，…，n。要注意與投擲銅板的情況不同，由於民調時樣本的抽取，屬於取出後不放回(總不能問同一人兩次)，因此各樣本間並不獨立，於是中央極限定理本來不適用。但因執行民調時，通常樣本數(1 千個左右)與選民數(通常總有幾萬人以上，否則便沒太大必要抽樣了)相比，往往不太大，因而將樣本取出後不放回，視同取出後放回(如此樣本便獨立了)，比起民調時由於拒答、未誠實回答，或問卷設計不當等，所產生的誤差，樣本間不獨立的誤差，其實尚可忍受。民調時通常設定信賴區間之信心水準為 95%，且預設估計誤差為 3%，由此得到預定的成功樣本為 1,068。經抽樣後，成功樣本數 n 很少會剛好等於 1,068。將實際的 n，代入誤差公式

$$d=0.98/n^{1/2},$$

便得此次抽樣的估計誤差 d。例如，若 n=1,100，則 $d\approx0.0295$；若 n=952，則 $d\approx0.0318$。

估計誤差 d 何以預設為 3%，誤差 1%不是更精準嗎？由誤差公式即知，d 若取為 1%，將使樣本數 n 從 1,068 增為

9,604，差不多是 9 倍了。民調往往有即時性，9 倍的樣本，將耗費更大的人力、物力及財力，也會讓調查時程更冗長，導致所得資訊失去時效。民調結果只是個參考，我們說過了，其中本已有相當多不易避免之各式各樣的誤差，與其試圖大幅降低 d，因而增加調查困難，不如努力改進調查品質(包含問卷設計及訪員訓練等)。另外，因 $\Phi(2.576)\approx 0.995$，若信心水準想由 95%增為 99%，則誤差公式中的 0.98 得改為 1.288，若 d 要維持在 0.03，則 n 便須增至 1,844。樣本數增加不說，99%的信心水準，看起來信心滿滿，卻因調查過程必會包含的各種大小誤差，因而宣稱高達 99%的信心水準，反會令人感到心虛。故也不認為有此必要。

一般而言，只要估計就會有誤差，95%的信心水準，表平均每得到 20 個信賴區間，其中就有 1 個不包含欲估計的參數。在實務上，若真能這樣便很好了。較令人擔心的反而是，調查品質不佳，使實際的信心水準，比 95%低很多。

附帶一提，設在某縣市抽樣，以估計對某候選人之支持度 p，得到成功的樣本數有 1 千餘，由此產生一信賴區間。由於各鄉(區)之成功樣本數，分別有數十個至上百個不等，因而又分別得到在各鄉(區)之 p 的信賴區間，並由誤差公式得到各鄉(區)之估計誤差。例如，若在某鄉(區)之成功樣本數有 64 個，則 d=12.25%。但這其實是有問題的，即使撇開其他因素不談，12.25%的估計誤差也未免太大了。

初期的大數法則及中央極限定理，均針對伯努力分佈，

此因這應是一最簡單又幾乎是最常見的分佈。生活裡的隨機現象，常出現恰有兩個結果的。如贏或輸、生男或生女，及格或不及格等。一試驗若恰有兩個可能的結果，不妨分別以成功及失敗稱之，若成功機率為 p，便稱此為一參數 p 之伯努力試驗(Bernoulli trial)。若以 X=1 或 0，分別表一伯努力試驗之結果為成功或失敗，則 X 有 $Ber(p)$分佈。對一數列 n 個獨立的參數 p 之伯努力試驗，我們常對其總成功數 S_n 有興趣，而 S_n 便有 $B(n,p)$分佈。S_n 的機率值中有階乘數，如 $n!=n(n-1)(n-2)\cdots 1$，相較於次方，如 a^n，$a\neq 0$，不論 n 多大，只要有個小計算器，經由取對數，很快便能知其約略的值。但對 $n!$，當 n 稍大時，便不容易求出。如若問 100!為何？便無法回答(但可藉助 Stiring 公式知其所在範圍)。二項分佈 $B(n,p)$常出現，卻在 n 較大時，其機率值便不好計算，此不便是當初發展中央極限定理的主要動機。又，$N(0,1)$分佈之機率密度函數為

$$\phi(x) = \exp(-x^2/2)/(2\pi)^{1/2}，x\in \mathrm{R}。$$

在 x-y 座標平面上，其圖形為鐘形，最高點在 x=0，左右對稱於 y 軸，向兩側下降，以 x 軸為極限。

由中央極限定理，若 S_n 有二項分佈 $B(n,p)$，則當 n 夠大時，標準化後的 S_n，即$(S_n-np)/(np(1-p))^{1/2}$，有近似的 $N(0,1)$分佈。只是所謂 n 夠大又是多大呢？不少教科書說，通常若 np 及 $n(1-p)$皆大於或等於 5，則以常態分佈來近似二項分佈 $B(n,p)$，便將夠精確了。但若 p 未知時，如何知 np 及 $n(1-p)$

是否大於 5？若有些資訊，如知 p 介於 0.4 與 0.5 之間，則 n 只要小小的 13 即可。但若無任何 p 之資訊呢？有另一指引。對一般期望值及變異數皆存在之一數列獨立且有共同分佈的隨機變數之和 S_n，很多教科書指出，“通常” n 若達到 30，則中央極限定理便適用了。30 個樣本不是很容易嗎？且還什麼分佈都可以呢！但實務上，是存在“不通常”的情況。如若隨機樣本之共同分佈，為參數 0.001 的波松分佈(Poisson distribution)$P(0.001)$，則樣本和 S_n 有 $P(0.001n)$分佈。即使 n 高達 500，S_{500} 有 $P(0.5)$分佈，並不必去近似。若硬要以常態分佈去近似會如何？就是誤差很大而已！對本例，恐怕要到 n 至少 7,000，才適合以常態近似。所以，實際應用時，仍得謹慎些。波松分佈取值在非負整數 0，1，2，…。在此對 $P(\lambda)$ 分佈，$\lambda>0$，其機率密度函數為，此分佈之機率值很容易求出，

$$f(k) = e^{-\lambda}\lambda^k/k!，k=0，1，2，\cdots。$$

關於隨機法則，我們有另一相當有用的法則，也是針對伯努力試驗。當 S_n 有 $B(n,p_n)$分佈，且 $n\to\infty$時，$np_n\to\lambda$，其中 $\lambda>0$ 為一常數，則 $n\to\infty$時，$B(n,p_n)$的分佈函數會收斂至 $P(\lambda)$的分佈函數。這便是稀有事件法則(law of rare events)。由此法則，在實務上，對某一發生機率 p 很小的稀有事件，若重覆觀測數 n 夠大，且 np“適當地”大時，則該事件共發生的次數 S_n(有 $B(n,p)$分佈)，便可以 $P(np)$分佈來近似。這種情況很多。若某地每天會發生因車禍致命的事件之機率為 0.004，則因 365×0.004=1.46，一年 365 天，會有致命車禍的

日數，便可以 $P(1.46)$分佈來當近似的模式。至於所謂 np 要適當地大，到底多大才屬“適當”？有些書說 n，p 宜滿足 $np\leq7$。那如果 $np>7$ 會如何？由於 p 很小，因而 $1-p$ 很接近 1，故此時亦必有 $n(1-p)>7$，如此中央定理便適用了。

稀有事件法則用途廣泛，諸如某保險公司每月意外死亡的理賠數、某醫院每天子時(晚上 11 點至 1 點間)進急診室的患者人數，及全世界每年墜機次數等，皆可以波松分佈當模式。每架飛機一次航行會墜落乃屬稀有事件，但全世界每年飛機航行次數相當多，因而全年的墜機次數，便有可能適合以波松分佈來當模式。波松分佈有個參數，此參數該如何得知？由於波松分佈之參數即期望值，故只要收集夠多的樣本，算出樣本平均，便可以之來估計參數。

我們說過，中外歷史上，曾實際去投擲銅板多次的人，應不會有太多。不過我國約在 1 千年前，便曾有 1 次投擲 100 個銅板者。在宋朝蔡絛著的“鐵圍山叢談”(宋朝史料筆記中，相當重要的一本)一書第二卷中，記載驍勇善戰的狄青(1008−1057)之一則事蹟。在率軍征討反宋的儂智高(1025−1055，一位領地在今越南的部族首領)前，大軍才一離開桂林，為鼓舞士氣，狄青“取百錢自持，與神約，果大捷，則投此錢盡錢面也期。”在萬人注視下，他揮手一擲，“百錢旨面”(旨面可能指有字那面)。狄青叫人將錢封住，說“俟凱旋，當酬神取錢”。待打勝仗回來，在眾人同觀下，他打開封條，“乃兩面錢也”。在民間信仰裡，狄青乃武曲星下凡(見“水滸傳”之“楔子”)，是宋朝傳奇名將，因而此事

蹟亦出現在一些通俗的兒童讀物中，讓狄青擲百錢一事，廣為流傳。事實上，投擲 100 個公正的銅扳，全出現正面，其機率為 0.5^{100}，約為 7.88861×10^{-31}。這麼小的機率，應沒人反對屬“不可能”會發生的事件。蔡絛是誰呢？有人說不定好奇。他生卒年不明，惟其父蔡京(1047−1126)倒是頗為人熟知，但並不是因他曾先後 4 任北宋宰相，而是因他乃“水滸傳”裡四大奸臣之一，且留下一著名的故事。第十五回的回目是“楊志押送金銀擔，吳用智取生辰綱”，其中生辰綱在唐宋時，指成批運送的生日禮物。在此回中，蔡京女婿北京大名府知府梁世傑，為賀他生日所送的“十萬貫金珠寶貝”，被書生吳用等人以智奪取。身為軍師，卻被取名吳用，“水滸傳”之作者，可能有“百無一用是書生”之意。

歷史上擲百錢者尚有一位，也是武將。在二月河(本名凌解放，1945−2018)的“康熙大帝”第三卷“玉宇呈祥”(下)裡，講了下述故事。清朝施琅(1621−1696)於攻打台灣前，拿出一把銅錢，向部屬說，“100 枚康熙銅錢，擲向台灣海域圖，倘若我軍出師順利，當有 95 枚以上的字面朝上。”一語既出，將台上下將士們無不失色。結果一擲之下，有 99 枚“康熙”字面朝上。當然立即軍心大振。原來施琅離京前，康熙皇帝特賜他 100 枚兩面皆字的銅錢，並告訴他如何操作。施琅怕有精明者起疑，特地換了 5 枚為正常銅板，如此反而讓屬下更相信此事為真。不過這故事，似未見於其他典籍，較可能是二月河仿之前狄青的擲幣所編出的，但做了些修改。因書是給今人看的，而現代人多少學過一些機率，

了解若 1 百個公正銅板皆出現正面，將無人會信其中沒有作假。只是出現 95 個正面就容易嗎？對投擲 100 個公正的銅扳，至少出現 95 個正面，利用二項分佈的機率密度函數，可輕易求出機率為 $79{,}375{,}496\times0.5^{100}$，約為 6.26162×10^{-23}。此機率是比出現 100 個正面大很多，有 7 千 9 百多萬倍。但仍是相當小，相較於 1 張統一發票，中特別獎之機率為 10^{-8}，此仍屬“不可能”會發生的事件。

大數法則及中央極限定理，有時會被混為一談。另外，偶會被過度推崇，以為能無所不在。如台灣在 2014 年 11 月底，所舉行的縣市長選舉投票前，有位台北市長候選人，拋出以遴選委員會，來挑選首長的主張。組遴選委員會的概念本來還好，但說到這裡就該結束了。只是世上懂得何時結束的人少，愛畫蛇添足的人多。這位市長候選人，又給了下述解釋，反而自曝其短：

> 在統計學上，N＞25，就會接近大數法則，也就是中央極限理論，不太容易出現偏頗的情況。雖然他準備設置的遴選委員會成員不到 25 人。但以過去經驗來看，只要 15 個人就會蠻準確的。

其中 N 是什麼並沒說明，應是指遴選委員的人數，且 N＞25 時，究竟何者會“接近”大數法則也沒說，較可能是指大數法則能“適用”。又大數法則與中央極限定理的內涵，乃完全不同，所以並無“大數法則也就是中央極限定理”的講法。且此二隨機法則，均針對隨機樣本(即獨立且有共同分

佈，惟此條件可以寬鬆些)。如今遴選委員會的成員，不外市長或各界推薦，與隨機樣本相差十萬八千里。不要說首長的遴選，與大數法則或中央極限定理均沾不上邊，就算遴選委員是隨機產生，且令 $X_i=1$，或 0，分別表第 i 個遴選委員對某特定首長候選人支持或不支持，$i=1$，2，…，N，則大數法則是說，N 夠大時(只是遴選委員的人數 N 當然不會有多大)，樣本平均 S_N/N 做為台北市民對某特定首長候選人支持率 p 之估計，不致於太離譜。至於支持率最高，是否便是最適合的首長人選？大數法則可沒說，且這也非大數法則之功能。至於中央極限定理乃用來求 p 之信賴區間，在這裡根本用不上。因而雖言之鑿鑿，且宣稱“蠻準確的”，我們卻必須說，若選出的首長人選不適合，絕不能怪大數法則或中央極限定理，否則便是甩鍋了。

9 隨機法則(三)

先來看一個例子。假設某袋中有 100 個球，其中有 30 個紅球，70 個非紅球。每次隨機取 1 球(即設袋中每個球被取中的機率皆相同)，每次取出後不放回。令 X_i=1，或 0，分別表第 i 次取中紅球或非紅球。經 n 次取樣後，得到一數列的伯努力隨機變數 X_1，X_2，…，X_n。又令 $S_n=X_1+X_2+\cdots+X_n$，表 n 次取樣後之樣本和，也就是共得之紅球數。首先我們注意到，此處 n 最大只能到 100，且 $S_{100}=30$，為一常數，因此 S_{100} 並無二項分佈。對取樣不放回，樣本數 n 是無法趨近至∞的。又，X_1 顯然有 $Ber(0.3)$分佈。那 X_2 之分佈呢？若 $X_1=1$(機率為 30/100)，則 $X_2=1$ 之機率為 29/99；若 $X_1=0$(機率為 70/100)，則 $X_2=1$ 之機率為 30/99。由此知 X_2 之值與 X_1 之值有關，因而 X_1 與 X_2 不獨立。又由上討論得

$$P(X_2=1)=30/100\times29/99+70/100\times30/99=0.3。$$

故 X_2 與 X_1 一樣，有 $Ber(0.3)$分佈。但 X_2 的分佈受 X_1 影響，故 X_1 與 X_2 並不獨立。另外，

$$P(S_2=0)=P(X_1=X_2=0)=70/100\times69/99\neq \mathrm{C}(2,0)\times0.7^2。$$

故 S_2 並無二項分佈。雖只驗證 X_1 與 X_2 不獨立，但直觀上，對 X_1，X_2，…，X_n，$1\leq n\leq100$，每一 X_i，$1\leq i\leq n$，皆有 $Ber(0.3)$ 分佈、彼此間並不獨立，且 S_n 並無二項分佈，證明在此略去。由此亦可得知，在摸彩活動中，先抽後抽並不影響各人的中

獎機率。

底下我們將上述例子一般化。假設一袋中有 N 個球，其中有 D 個紅球，及 N-D 個非紅球。自袋中依序隨機取出 n 個球，$1\leq n\leq N$，每次取出的球皆不放回。令 X 表總共取出之紅球數。則

(6) $P(X=k)=\mathrm{C}(D,k)\mathrm{C}(N-D,n-k)/\mathrm{C}(N,n)$，

$$\max\{0,n-N+D\}\leq k\leq\min\{n,D\}，$$

其中 min 及 max，分別表較小及較大者。至於 X 可能取的值 k，何以有那麼複雜的範圍？此因取 n 個球，且袋中紅球數共有 D 個，所以取中的紅球數 k，既不會超過 n，也不會超過 D，即 $k\leq\min\{n,D\}$；另一方面，共取出的非紅球數，即 n-k，當然不能超過全部的非紅球數 N-D，且 $k\geq 0$，故 $k\geq\max\{0,n-N+D\}$。(6) 式便定義出一超幾何分佈 (hypergeometric distribution)，有 3 個參數，此分佈以 $H(N,D,n)$ 表之。前述 S_n，便有 $H(100,30,n)$分佈。民調裡的取樣，大抵都屬取樣不放回的情況，因而此時涉及的，便是超幾何分佈，而非二項分佈。

取樣不放回的例子處處可見。除了民調外，品質管制裡也常出現。如要檢驗一批電池的壽命是否合格，先隨機取樣，由於取出的每一個都要測試，所以當然是取樣不放回。另外，對台灣某保育的野生動物，如被列入“其他應予保育”的台灣長鬃山羊(學名 Naemorhedus swinhoei)，如何估計

在某棲息地區，共有多少頭？一種容易想到的作法是，設該地區此種山羊有 n 頭，在該地區捕捉 k 頭後，做好記號後便放走，隔若干日後再度捕捉 t 隻，算出其中有做記號的有 s 隻。假設每頭山羊被捕捉到的機率都一樣，則再度捕捉的 t 隻中，有做記號的山羊之比例，與全部中有做記號的比例應差不多。令 $k/n=s/t$，解出 $n=kt/s$。我們便以 kt/s，來估計該地區的山羊數。像這類對動物的捕捉－再捕捉法 (capture-recapture method)，通常亦是取樣不放回。惟此作法有些缺點，首先是記號可能脫落，其次是不知動物是否每隻被捕捉到的機率皆相同？若差異很大，則此估計法便將有不小的誤差。但在缺乏更有效的辦法下，此仍是一常見之估計法。由上述說明可知，對於取樣，就是會存在各式各樣的誤差。因而引用中央極限定理時，樣本獨立且有共同分佈之條件，實務上並不易滿足。但即使如此，取樣時仍須盡量謹慎，使樣本儘量滿足或近似滿足，獨立且有共同分佈的條件。

假設某地有 N 個合格選民，其中有 D 個人支持某候選人 A，有 N-D 個人不支持 A。欲估計 A 之支持率 $p=D/N$。隨機抽取 n 個選民調查，$1\leq n\leq N$，每次取出皆不放回。令 X 表其中支持 A 之總人數，則 X 有 $H(N,D,n)$分佈。要注意的是，N 應還能知道，但 D 並不知，否則 $p=D/N$ 便知道了，何須估計？由於取樣不放回，因而 X 並無二項分佈，所以原本無法利用中央極限定理來近似。但我們之所以仍引用，是基於忽略取樣不放回所造成的誤差。例如，比較第一及第二次，被抽中選民支持 A 之機率，若第一次抽中之樣本不屬於 D，則

$$D/N - D/(N-1) = -D/(N(N-1))，$$

若第一次抽中之樣本屬於 D，則

$$D/N - (D-1)/(N-1) = (N-D)/(N(N-1))，$$

不論那一情況，當 N 很大時，二機率之差便都很小。由此即得知，只要抽出的樣本數 n 與 N 相比很小，則若將取樣不放回，當做取樣放回看待，產生的誤差並不太大，尚可忍受。這是取樣不放回時，樣本裡出自 D 的個數，可近似地以二項分佈來描述，進而引用中央極限定理，以常態分佈來近似二項分佈的原因。換句話說，一路下來，其實誤差不少，因而我們已數次強調，民調的結果，包括所得支持率 p 之信賴區間，都是僅供參考，不能太當真。

有些人，秉於實事求事的精神，企圖證明 $N\to\infty$時，

$$\mathrm{C}(D,k)\mathrm{C}(N-D,n-k)/\mathrm{C}(N,n)\to\mathrm{C}(n,k)(D/N)^k(1-D/N)^{n-k}，$$

亦即 $N\to\infty$時，超幾何分佈 $H(N,D,n)$分佈趨近至二項分佈 $B(n,D/N)$。可惜其證明過程不必看便知是錯的。因既然令 $N\to\infty$，則其極限下的結果，怎能還包含 N 呢？事實上，如上段之說明，以常態分佈來近似超幾何分佈，就是一關又一關地忍受誤差，而根本未讓任何參數趨近至∞。這是統計實務裡常見的作法，就是對於一個又一個“不太大”的誤差，忍受再忍受。

“大數法則”4 字淺顯易懂，因而經常出現，但有時卻

是與中央極限定理混淆。此外，有時明明是稀有事件法則，也會被誤以為是大數法則。另有一巨數法則(law of truly large numbers)，與大數法則無關，但在“凡間”，也常大剌剌地以大數法則自居。其英文稱呼中的 truly large numbers，更如鳩佔鵲巢。此處的“鵲”，當然便是大數法則了。巨數法則通常出現在科普文章中，有時甚至就稱為大數法則(law of large numbers)。此法則是說，“當樣本數夠大，任何非比尋常的事(any outrageous thing)，都可能發生”。非比尋常，或聳人聽聞等，皆屬不可能發生的事件，也就是機率極小的事件。巨數法則指出，不論發生機率再怎麼小的事件，一旦有夠多的樣本，其發生便不足為奇了。例如，每年過年期間，台灣有些廟宇會舉行擲筊比賽，連得最多聖筊(一陽一陰)者，可獲汽車之類的優渥獎品。每得 1 次聖筊之機率為 1/2，有可能連得 20 次聖筊嗎？$(1/2)^{20}=1/1,048,576$，小於百萬分之 1 的機率，算是相當不容易。但台灣有約 2 千 3 百萬人，即使扣除約 1 百萬不滿 5 歲之幼兒，仍有約 2 千 2 百萬人，若每人都去擲筊，將會見到 21 件左右至少連得 20 次聖筊者。至於全世界人口超過 70 億，只要能每人都去執行，則即使發生機率約 10 億分之 1 的事件(如連得 30 次聖筊，其機率約為 9.31×10^{-10})，其發生便一點都不稀奇。所謂天下之大，無奇不有，正是這個意思。

小機率事件發生，向來不可輕易忽視。會出現在媒體上的報導，便有不小比例，屬於小機率事件，因小機率才引人注目。某君去家網路上評價不錯的餐廳用餐，結果某道菜不

新鮮，經理誠懇道歉，並給適當補償，此君便不計較了，接受只是自己運氣不佳。若隔幾個月再去，又遇到食物不新鮮，則該君可能便再也不去了。因他已不願相信純粹是自己運氣不佳，而認為該餐廳品管有問題。這就是小機率事件的影響力。中文裡的“三人成虎”，及“曾參殺人”等典故，都是在強調小機率事件的影響力。底下給一小機率事件備受矚目之例。

有位大學生，在某年暑假 7、8 月間，與同學常在網咖打發時間，得到百餘張發票。開獎後中了 12 張，幸運之神眷顧，該生高興萬分。豈料兌獎後，他被國稅局通知去說明，有人還建議他，乾脆將全都獎金都繳回算了，省得麻煩。只不過中最小的六獎，獎金區區 200 元，全部才 2,400 元，連想好好請個客都不容易，便被懷疑其中有弊，運氣好有罪嗎？不但該生，也有不少人感到不服氣。

六獎是發票號碼末 3 位，與頭獎中獎號碼(有 3 組)末 3 位相同，那一期(兩個月一期)增開 2 組六獎，中獎機率提高至 0.005(=1/200)。五獎(獎金 1,000 元)是發票號碼末 4 位，與獎中獎號碼末 4 位相同，中獎機率 3/10,000，四獎以上當然中獎更難。一般人並不妄想，因此所謂中獎，不特別說明時，就是指中六獎，且通常中 1 張就很滿意了。新聞上報後，有自認數學不錯者提出其算法，“連中 12 張”之機率為

$$(1/200)^{12} = (1/4{,}096)\times 10^{-24} \approx 2.44\times 10^{-28}，$$

以此佐證該生中獎必有弊。此機率固然微乎其微，但卻是錯

的。因何須“連中”呢？此生被懷疑，是因他擁有的發票中，有 12 張中獎，並不必連中，正確的機率應比連中大很多。

運氣好沒有不行，但過度好的運氣，讓人產生懷疑，也是合理。國稅局負責發票中獎的官員，總不能尸位素餐，見可疑是該追查一下。媒體未明確報導該生究竟幾張發票，只含糊地說百餘張，我們就以 150 張計。令 X 表中獎張數。由於同一家店開出的發票，應為連號，故各張發票的中獎與否，不會獨立，因此 X 並無二項分佈。但我們只是想約略了解中獎 12 張，究竟有多不可能，不妨將 X 之分佈，就視為 $B(150,0.005)$。則 X 之期望值與標準差，分別為 $150\times0.005=0.75$，及 $(150\times0.005\times0.995)^{1/2}\approx0.8639$。$X\geq12$，表 X 超過期望值約 13.02($\approx$(12-0.75)/0.8639)個標準差，不論對那一分佈，此機率想必都是很小的。附帶一提，因 $150\times0.005=0.75<5$，故此處不宜引用中央極限定理來近似二項分佈。但可利用稀有事件法則，得 X 有近似的 $P(0.75)$分佈。對 $P(0.75)$分佈，$X\geq12$ 的機率當然很小。事實上，若 X 有 $B(150,0.005)$分佈，利用計算機，可求出 $X\geq12$ 之機率約為 2.2211×10^{-11}。對發生機率這麼小(比連得 35 次聖筊的機率約 2.910×10^{-11} 還小)的事件，的確不能以一句“運氣好”，就想輕鬆帶過。

10 中位數

北美有四大職業運動聯盟，除 NBA 外，尚有 NFL (National Football League，國家美式足球聯盟)、MLB(Major League Baseball，美國職業棒球大聯盟)，及 NHL(National Hockey League，國家冰球聯盟)。若依聯盟的收入及影響力等來評比，NBA 在四大聯盟裡，大約會落到第三。不過，若論球員薪資，NBA 卻是排名第一。只是 NBA 有 4、5 百個球員，如何呈現球員之薪資呢？數據的代表值，最先想到的很可能是平均。但平均是球員薪資之恰當的代表值嗎？

2020 年 12 月 17 日，有媒體報導，“依網站 Statista 的資料，四大聯盟中，今年(2020)賽季的球員平均年薪，以 NBA 球員的 832 萬美元居首，而且是平均年薪第二名 MLB 的 403 萬美元之兩倍以上。”雖被公佈平均年薪高達 832 萬美元(以今日匯率，約 2.3 億台幣)，但大多數的 NBA 球員，很可能會抱怨，大眾被平均誤導了，他們的收入遠比平均低多了。是這樣嗎？沒錯！同年 4 月 18 日有另一則報導，其中提到，“NBA 現役有 450 位球員，…。2017 年 NBA 球員平均年薪是 693.6154 萬美元，年薪中位數(median)是 250 萬美元，2019−2020 球季的最低年薪則是 89.331 萬美元，低收入球員和頂薪球員間有極大差距。”報導中透露年薪中位數才約平均年薪的 36.04%。再前一年(2019 年)的 1 月 22 日，則有如下報導，“NBA 再一次成為全世界球員平均年薪最高的聯

賽，達到 777.1 萬美元；年薪中位數為 432 萬美元。”在此則報導裡，年薪中位數沒低那麼多，但也才約平均年薪的 55.59%。我們再給一 2018 年 2 月 24 日的報導，“2017−2018 賽季，NBA 在球隊領薪水的共計有 548 人次(包括少數拿多份工資的)。而本賽季，NBA 總薪資為 33.5 億美元，去掉短工(收入 30 萬以下的)，還有 495 名球員，人均薪水 677 萬美元…。我們看中位數，其中薪資排第 248 名的球員年薪 329 萬美元(拿同樣年薪的共計有 4 名球員)。”其中指出球員薪資的中位數，約為平均年薪的 48.60%。又第 248 名剛好居全部 495 名的正中間，即說明中位數，乃排名“居中那位”。

也許年份及計算的基準不同，因連 30 支球隊的球員總數，都有 548、495，及 450 等數字，上下差了將近 1 百人，導致前述數筆 NBA 球員年薪之平均及中位數，有不小差異，便不足為奇了。但 3 處提到年薪中位數，皆比平均年薪少很多，更有兩處的中位數連平均之半都不到。不難理解，這主要是因有些明星球員，年薪高達 4、5 千萬美元(甚至年薪還不是其主要收入，更多收入來自廣告代言等場外收入)，一支球隊中，兩個大牌球員的年薪(其高薪即所謂極端值)，常便超過全隊其他球員的年薪和，遂將球隊整體平均年薪大幅拉高。因而對職業運動，平均年薪往往不是球員薪資之最恰當的代表值，中位數才是。求一筆數據之中位數，看起來很簡單，只要將數據排序即得，不像平均，還要先求總和。

中位數常可消除極端值的影響。例如，對 1，2，3，4，5，平均及中位數皆為 3；但對 1，2，3，4，5,000，平均為

1,002，那一特別大的數 5,000，讓平均數增大不少，但未更動中位數，仍為 3。

再來看一則國內的報導。2020 年 12 月 23 日，行政院主計總處發布標題為“108 年工業及服務業受僱員工全年總薪資中位數及分布統計結果”之一則新聞，其中提到，“2019 年工業及服務業受僱員工全年總薪資中位數為 49.8 萬元，…。薪資中位數是將員工依薪水由低至高排序，位處中間者的薪資即為薪資中位數，因此相較平均數容易受極端值影響，中位數更能反映多數民眾的薪資狀況。…，2019 年全體工業及服務業受僱員工數為 796.7 萬人，薪資中位數為 49.8 萬元，意即有半數受雇員工年薪未達此水準，約 398 萬人。”其中說明為何採中位數，及中位數乃“位處中間”那數。

生活裡，人們求平均的機會多，求中位數的機會少，但直觀上，中位數就是正中間那數，倒不是困難的概念。學生時代，自小學起，便不時會按高矮排隊型，那時並未太在乎精準度，以目測差不多即可。身高居中的是那個？如果 9 個人，就正中間那位，即第 5 個的人之身高。如果 10 個人呢？沒有唯一的中間，那就中間兩位，即居第 5 及第 6 的兩個身高，都可算居中，反正差不了太多。若只能給一值呢？便取中間兩人身高之平均，還是同一句話，反正差不了太多。一般而言，設數據個數為 n，將數據按小至大排序後，當 n 為奇數時，令第$(n+1)/2$ 個值為中位數；當 n 為偶數時，則第 $n/2$，及$(n+1)/2$ 個，皆可為中位數，或者取第 $n/2$，及$(n+1)/2$ 個之平均為中位數。當 n 為偶數時，若依第一個定義，中位

數為數據中的某兩個；若依第二個定義，中位數乃唯一，但並不屬數據中的任一個。

其實人們對中位數如何產生？通常並不太計較。在求 NBA 球員年薪的中位數時，往往連球員總數，都搞不定，上下可差到約百人，因而正中間究竟是那位，並不易確定。另外，因受傷而整年沒上場的球員，其薪資又如何採計？於是在看到報導中的 NBA 球員年薪之中位數時，很少人會去過問計算基準。反正只是個參考，想對球員年薪，有個約略的概念而已。此正如有些人在談到自己年薪時，會思考一陣子：加班費及績效獎金等算不算？每月被扣除的午餐、交通車及退撫儲金等費用，要不要加回來？連自己年薪究竟有多少，都要想一想該講那個值，因而對所見到之各種中位數，便未太追究其涵意。即使如此，“居中那位”，或“位處中間”之概念，大家還是有的。

生活中離開不了統計，媒體上經常可看到各種統計數據的報導，其中不乏百分位數(percentile)及中位數等。百分位、中位，字面上看，令人覺得淺顯易懂，於是便出現在國中數學了，並歸入統計領域。先看“中位數”？在之前，2008 年通過，且自 2010 年開始實施的“9 年課綱”(九年一貫數學課綱)裡，中位數放在國三。在“細目銓釋”(分年細目銓釋)裡有，“能認識平均數、中位數與眾數；能認識全距及四分位距，並製作盒狀圖；能認識百分位數的概念，並認識第 10、25、50、75、90 百分位數”。對於中位數，於“分年細目銓釋”裡，捨“居中那位”及“位處中間”，相當口語地說，

“中位數是將資料排序後，前後各切一半的中間位置資料值。…中位數會使落在兩邊的資料呈現出某種平衡狀態。…中位數則是個數的平衡。”“呈現出某種平衡狀態”，有點文青的寫法，但個數究竟如何平衡？並沒明說。這就算了，依此定義求中位數，卻時感困惑。例如，對於數據個數為偶數，如 1，2，3，4，5，6，前後各切一半，有一半是 1，2，3，另一半是 4，5，6，中間位置 3.5，是這樣嗎？因提到要有平衡狀態，那看來就是 3.5 了。但對於數據個數為奇數，如 1，2，3，4，5，中間位置顯然為 3，但此時如何前後各切一半？將 3 這個數字劈為兩半嗎？這種“所羅門的審判”(The Judgement of Solomon，典故見“舊約聖經”“列王紀上”第 3 章第 16-28 節)式的用語，看似豪邁，卻只有數據個數為偶數時，中位數才求得出。不得不說此定義並不周詳。但定義不妥的，並不僅止於此。

中位數為百分位數之一種，即第 50 百分位數。在“9 年課綱”的“用詞解釋”(標準用詞與解釋)裡，便藉百分位數來引進中位數。即“第 50 百分位數，通常表示比這筆或這組數大和比這筆或這組數小的資料各佔一半”。依此定義，我們來重新檢視前述 2 筆數據。對 1，2，3，4，5，顯然 3 不是中位數了，因比 3 大及比 3 小，各有 2 個數，並不佔 5 個數之半，即 2.5 個。此顯示當數據個數為奇數時，依“用詞解釋”，根本不會有中位數存在。其次對 1，2，3，4，5，6，顯然區間(3，4)中的任一數皆為中位數，但端點 3 與 4 則皆不是。也就是數據個數為奇數時，中位數並不存在；數據

個數為偶數時，中位數不但不唯一、有無限多個，且沒有一個在數據中。中文裡，“銓釋”與“解釋”不知有何差別，但在同一份課綱的“分年細目銓釋”及“用詞解釋”裡，給出迥然不同的中位數之定義，且二者在數據個數為奇數時皆窒礙難行，未免太奇怪。

幸好，在教育部主編，2012 年 12 月出品，針對“9 年課綱”的“補救教材”(國中數學基本學習內容補救教材)中說，“甚麼是‘中位數’呢？通常對於給定的一組資料，將資料的數值由小到大排列，(1)若資料個數是奇數個，則最中央的一個資料數值是中位數，(2)若資料個數是偶數個，則最中央的兩個資料數值的平均值是中位數”。當然將資料的數值由大到小排列亦可。今日國中，大致便採此中位數唯一之定義。依此，只要給定一組數據，可明確求出其中位數。這的確是份“補救”教材，因中位數的定義，終於弄清楚了。這樣講實在有點難堪，因中位數不是本就該如此定義嗎？在媒體的報導裡，中位數不一向就是這麼解釋的嗎？最多就是當數據有偶數個時，中位數須唯一或允許可以有兩個之別。如今大費周章編出來的課綱，卻需要補救？依“補救教材”，即使數據裡有重覆的，仍可找出中位數。如對數據 1，1，1，1，2，中位數為 1；對 1，1，1，2，2，2，中位數為 1 與 2 之平均，即 1.5。

底下再給一常見的中位數之定義。對一組數據，若在某值之前及之後，皆至少有 50%個，則該值便為此組數據之中位數。如一般用法，這裡的“之前”及“之後”，都表包含

該值。依此定義，對 1，2，3，4，5，顯然 3 為中位數；而對 1，2，3，4，5，6，顯然 3 及 4 皆為中位數。甚至若允許中位數不在該組數據中，則區間[3，4]中的每個數，便皆為中位數。此定義可行，當數據個數為奇數時，中位數唯一；當數據個數為偶數時，中位數有無限多個。

繼“9 年課綱”後，2019 年開始實施的“12 年課綱”(十二年國教課綱，又稱 108 課綱)，數學在國中 1 年級有，“統計數據：用平均數、中位數與眾數描述一組資料的特性；使用計算機的 M+或∑鍵計算平均數”。中位數的定義，在這份新課綱裡有釐清楚嗎？

課綱向來寫得簡略，常連微言大義都稱不上。在針對“12 年課綱”完成的補充材料“課程手冊”(數學領域課程手冊，2020 年 2 月，國家教育研究院發行，共 765 頁)裡，再度出現“中位數是將資料排序後，前後各切成一半的中間位置資料值”。此顯示編寫這份“課程手冊”者，仍喜愛各切成一半的豪邁講法，當能也可能是照抄舊課綱裡的資料。我們已說明了，依此定義，當數據個數為奇數時，中位數不存在。幸好，“12 年課綱”公佈後，國中數學課本的編輯，知道不必理會“課程手冊”，守住“9 年課綱”時代，“補救教材”裡中位數之定義，而不讓中位數困惑師生。數學課綱裡的中位數，不知要到那一年的版本，才能停止需要補救。

有些數學家不習於動手，十年來，不但未曾依自己給的，各切成一半之定義去求中位數，對媒體上不時出現的報

導裡，特地附上中位數的意義，可能也未曾仔細看。原本“9年課綱”國中裡的中位數及百分位數，長期引起的學習困難，解決之道，居然僅是在“12 年課綱”裡，保留中位數在國中，而將百分位數移至高中，但對中位數的講法卻維持不變，頗有鋸箭法之風格。高中裡的百分位數，會被寫成什麼樣子，不禁令人忐忑不安。

11 百分位數

行政院主計總處，於 2020 年 12 月 23 日，發布一則標題為“108 年工業及服務業受僱員工全年總薪資中位數及分布統計結果”之新聞，其中提到，“2019 年工業及服務業全年總薪資第 1 十分位數為 29.1 萬元，較 107 年增加 4.10%；第 9 十分位數為 117.9 萬元，則增 2.62%”。其中第 1 十分位數，即第 10 百分位數，第 9 十分位數，即第 90 百分位數。在有關薪資的報導，有時除中位數(第 50 百分位數)外，會更細緻地給幾個特別的百分位數，以讓人們更加了解薪資結構。有助解讀數據，百分位數本就讓一般人不會覺得太陌生。過去二十餘年，由於學測(大學入學學科能力測驗)的頂標、前標、均標、後標，及底標等 5 標，乃分別依到考生之第 88、75、50、25、12 等百分位數之級分而定，及基測(國民中學學生基本學力測驗，2013 年舉行最後一屆，之後由會考(國中教育會考)取代)之 PR(percentile rank，百分等級)值，讓人們更常接觸到百分位數。

或許是因太常在媒體出現，所以百分位數，被認為是國民必須懂的統計概念。在之前的“9 年課綱”裡，置於國中 3 年級。經過多年，可能是因被認為較簡單，在“12 年課綱”裡，中位數由國中 3 年級移至 1 年級；至於百分位數，可能被認為較難吧，則移至高中 1 年級，列在“數據分析”項下。於“學習內容”(學習內容條目及說明)中寫著，“一維數據

的平均數、標準差。二維數據的散布圖，最適直線與相關係數，數據的標準化。”就這樣而已？並沒出現百分位數一詞啊！這是課綱的習性，究竟要學些什麼，不能只看“學習內容”。學習的內容怎能不在“學習內容”裡？對課綱並無法要求邏輯，即使是數學課綱。在該項之“備註”裡有，“適度與國中所習的數據分布圖重疊，但加深加廣其情境，並將四分位數延伸至百分位數。學生應知道統計數據可能有略為不同的定義，也應理解可能產生數值略為不同但意義相同的數據；…。”“學習內容”裡，只提到平均數及標準差，然後在“備註”中，憑空冒出四分位數，且說要延伸至百分位數。將部分該學習的內容，放在“備註”中，古人說綱舉目張，如今綱目不分了。至於那句“產生數值略為不同但意義相同的數據”，並無法理解。至於四分位數又是什麼？

“12 年課綱”的國中 3 年級，在“統計數據的分布”項下，於“學習內容”中寫著，“全距；四分位距；盒狀圖”。在該項之“備註”裡有，“…，本條目則傳達以盒狀圖描述數據的集中程度。”只是尚未引進百分位數，四分位數會容易定義嗎？令人存疑。我們列出“課程手冊”中，該項之基本說明的第(2)及第(4)點如下：

> (2). 認識第 1、第 2、第 3 四分位數(可記為 Q1、Q2、Q3)的意義，知道如何運用資料的累積相對次數分配表來找出 Q1、Q2、Q3。知道第 2 四分位數即為中位數；四分位距則為第 3 四分位數與第 1 四分位數的差，即 Q3-Q1。

(4). 關於四分位數的定義，一般並沒有統一的取法，很多處理資料的統計軟體所用的四分位數定義也不一樣。因此學生應先確切學習取四分位數的原理，再學習其計算法則。

第(4)點令人疑惑，含混地說以“累積相對次數分配表”找出，且說“統計軟體四分位數的定義也不一樣”，卻不給一個定義。另一方面，要“學生先懂四分位數的原理，再學其計算法則”。但原理是什麼？計算法則是什麼？就不說了，有故弄玄虛之嫌。雖然沒給定義，“課程手冊”有給一“釋例”。一班有 30 個學生之數學段考成績(我們將成績由小到大排列)如下：

5，33，35，35，36，43，44，45，47，48，52，55，58，64，64，65，68，69，70，74，75，78，78，79，80，82，83，84，85，89。

然後立即給出“最大數與最小數分別為 89 與 5，所以全距為 84。以及 Q1、Q2、Q3 分別為 45、64.5、78，則四分位距為 Q3-Q1=33”。至於 Q1、Q2 及 Q3 如何求出，卻不說。看起來“課程手冊”之編寫者，有可能不認為在引進百分位數之前，適合講此題材，乾脆只給答案，如何得到不講。附帶一提，Q1、Q2、Q3，可別分為第 25、50，及 75 百分位數。但國中數學裡，有什麼必要放進一無法明確給出定義的四分位數？突兀地講授四分位距及盒狀圖的目的何在？這些都令人不解。

在“9 年課綱”的“標準用詞與解釋”裡，說百分位數是，“各筆或各組資料的相對位置，表示有百分之多少的資料比該筆或該組資料的數要小”。這定義與一般的認知大不相同。依此定義，設有數據 1，2，…，100。則 51 是中位數，但 50 卻非中位數，相當奇怪。其他更不要說第 1 百分位數為 2，…，第 99 百分位數為 100 了。

前面說過，“12 年課綱”的高中 1 年級，在“數據分析”項的“備註”裡，說“將四分位數延伸至百分位數”，但其實沒什麼好延伸的，先教百分位數才較恰當。“課程手冊”裡，關於此部分寫得相當長。在“相關約定”的(1)裡寫著，“數據資料的第 m 百分位數記作 P_m，其中 m 為正整數 1，2，…，99。其詳細約定，寫在釋例中。”但不知為何，在“釋例”裡並無“例”，只說 P_m 指的是同時滿足以下兩條件的數，“小於或等於 P_m 的資料數量至少占全部資料量的 m%；大於或等於 P_m 的資料數量至少占全部資料量的 $(100-m)$%。當資料中恰有一個滿足上述條件的(原始)數據時，採用它作為 P_m；當超過一個(原始)數據滿足上述條件時，取它們的平均值作為 P_m。”至於四分位數，則連提都不提，所以由四分位數延伸至百分位數之議題，完全被拋開了。

依上述“課程手冊”之“釋例”，再度來看這組相當普通的數據 1，2，…，100。由第 1 個條件，立即得 P_1=1、2，P_2=2、3，…，P_{50}=50、51，…，P_{99}=99、100，再由第 2 個條件，得 P_1=1.5，P_2=2.5，…，P_{50}=50.5，…，P_{99}=99.5。雖依“釋例”，求出了百分位數，只是這樣求出的百分位數，

與一般人所想的很可能不同。既然是 1，2，⋯，100，大部分的人可能想當然耳地以為，P_1、P_2，⋯，P_{99}，分別是 1，2，⋯，99。很難想像會冒出 1.5，2.5，⋯，及 99.5 等百分在數。

換組數據來看。假設有數據 1，2，⋯，99，即將上筆數據去掉 1 個。得 $P_1=1$，$P_2=2$，$P_3=3$，⋯，$P_{50}=50$，⋯，$P_{99}=99$。有趣的是，數據少掉一個，百分位數便能皆如預期了

再看一筆數據 1，2，⋯，50。先得 $P_1=1$，$P_2=1$、2，$P_3=2$，$P_4=2$、3，⋯，$P_{50}=25$、26，⋯，$P_{98}=49$、50，$P_{99}=50$。當 m 是奇數時，P_m 只有一個值；當 m 是偶數時，P_m 有兩個值。由此得 $P_1=1$，$P_2=1.5$，$P_3=2$，$P_4=2.5$，⋯，$P_{50}=25.5$，⋯，$P_{98}=49.5$，$P_{99}=50$。除 P_1、P_{99}，及中位數外，其他百分位數，可能讓人有些糊塗吧！

對數據 1，2，3，4，5，先得 $P_1=P_2=\cdots=P_{19}=1$，$P_{20}=1$、2；$P_{21}=P_{22}=\cdots=P_{39}=2$，$P_{40}=2$、3；$P_{41}=P_{42}=\cdots=P_{59}=3$，$P_{60}=3$、4；$P_{61}=P_{62}=\cdots=P_{79}=4$，$P_{80}=4$、5；$P_{81}=P_{82}=\cdots=P_{99}=5$。由此即得 $P_{20}=1.5$，$P_{40}=2.5$，$P_{60}=3.5$，$P_{80}=4.5$，其餘不變。除 4 個外，其他百分位數，都是多個取同一值。

在“9 年課綱”的“細目銓釋”裡，於“能認識百分位數的概念”項下，有 5 點說明，其中第 2 點為，“知道百分位數通常用於分析總次數多的資料，避免在資料數少的例子中，做百分位數的教學。”上述幾個例子顯示，數據若太少，一目了然，本不需藉百分位數來解讀數據，此時若仍求出百分位數，不但不會讓數據更清楚，反而是讓人覺得百分位數

礙手礙腳。數據較多時會如何呢？底下給一之例。假設 NBA 球員有 450 位，按年薪由低至高排序，第 1-4 位，年薪各 20 萬(美元，底下同)，第 5 位，年薪 80 萬，第 447-450 位，年薪各 4.5 千萬，第 446 位年薪 3.5 千萬(美元)。則依上述“釋例”，可求出 P_1=80 萬，那 4 個 20 萬，對求 P_1 毫無影響；至於 P_{99}=3.5 千萬，那 4 個 4.5 千萬，對求 P_{99} 也毫無影響。於是人們對 NBA 球員年薪的理解，低薪被拉高，高薪則被拉低。至於 P_2，為第 9 與第 10 位之年薪，所以是二者之平均，P_3 為第 14 位之年薪，…。有很多球員的年薪被棄而不用，對百分位數沒有貢獻。但對沒學過中位數者，他可能會將最低 5 位的年薪平均，當做 P_1，第 6 位至第 9 的年薪平均，當做 P_2，…，第 446 位至第 450 位的年薪平均，當做 P_{99}，這樣說不定更符合百分位數的精神呢！由本例知，即使數據較多，百分位數也不見得能大展長才，有效讓人了解數據。也就是百分位數，不易有一能處處令人完全滿意的定義。

百分位數常用在薪資。政府每年會公佈前一年國民所得，也會列出一些 5 百分位數。但怎樣叫有所得？打工、租屋、擺地攤，及網路交易等，政府都能精確掌握嗎？究竟列入考慮的，是 900 萬人，或 1 千 1 萬人？不同單位所做的統計，差異可能很大。因此其中百分位數如何產生，太計較意義並不大。反正只是想讓人約略了解國民所得，高到那裡去，及低到那裡去。講起來就是件差不多之事，無法深究。這樣不求弄太清楚的題材，說起來，並不適合出現在講求精準之中學數學課程裡。

12 學測級分及 5 標

中學生或許不會太關心薪資的百分位數。有關成績的百分位數，他們倒是可能較常接觸到。除了校內成績外，學測的頂標、前標、均標、後標，及底標等 5 標，乃分別依到考生之第 88、75、50、25、12 等百分位數之級分而定。5 標既是由級分之百分位數得到，底下我們便先來看級分。

台灣的學測，乃自 1994 年開始實施，高中生想進大學，學測成績相當關鍵。一直到 2021 年，學測共有國文、英文、數學、社會及自然等 5 科。因應“12 年課綱”，自 2022 年起，數學將分成數學 A 及數學 B，總共 6 科。各科原始滿分不盡相同，國文、英文及數學皆為 100 分，自然 128 分，社會 144 分。但原始成績只是過客，精心設計的 128 及 144 都沒用，“大考中心”一律將每位考生的各科原始成績，皆轉換成 0 至 15 的“級分”。何以要全換算成 15 級分制？“大考中心”的說法是，“為了不要讓考生為了 0.1 分的差距而計較”。但這理由很可笑，一方面成績豈有不計較的？另一方面，原本 0.1 分的差距，通常會被理解為沒差，不太會計較，經轉換後，有可能變成 1 級分的差距，差距變大，那不是得計較了？至於原本差 5 分，那應真的有差距了吧！經轉換後，卻可能同一級分，想計較也由不得你。所以，級分的轉換，有些人得利，有些人不利，完全憑運氣，與能否讓考生不計較分數無關。本來考試除實力(必然性)外，已有些不

得已的運氣成分(隨機性)。“大考中心”便是將隨機性更加重。

級分如何得出？各科取原始成績前 1%考生(小數無條件進位至個位數。如若某科到考生有 128,424 人，1%是 1,284.24 人，則取原始成績最高的 1,285 人)之平均成績除以 15(以四捨五入取到小數第 2 位)，除後假設得到 a，a 即為級距。於是原始成績 0 分為 0 級分，介於(0，a]為 1 級分，介於(a，$2a$]為 2 級分…，介於($13a$，$14a$]為 14 級分，而($14a$，滿分]為 15 級分。要注意諸如(0，a]，並非實數上的區間，因成績只計算到小數第 2 位。所以(0，a]，乃表成績由 0.01，0.02，…，至 a。

自 2002 年度起，學測除了級分外，各科及總級分，又都依到考人數之百分位數，訂出頂標、前標、均標、後標，及底標等 5 項標準，此即所謂 5 標。學測成績，能用在諸如個人申請等，進入大學的管道。各大學校系在第一階段篩選時，可對學測的總級分，或各科級分，依 5 標訂出檢定標準，做為篩選門檻。台灣學生申請進美國大學就讀時，要附托福(TOEFL，Test of English as a Foreign Language)成績。托福測驗，自 2005 年起，採網路考試(iBT，Internet-Based Test)，總分為 120 分，成績有效期限為兩年。考過托福者，只須講成績如 80、92 或 104 等，不必說在何時或何地考。測驗能做到這樣的標準齊一，當然是很厲害的。學測呢？

“大考中心”有專家認為，“現在的學測成績採用浮動

15 級分法，因此其級分以及 5 標較不受試題難易之影響，每年均相當穩定”。雖對級分及 5 標之穩定性，講得相當肯定，只是至少在數學科，這並非事實。依原始成績等距來給 15 級分，絕對不表每一級分的考生數，約有 6.6%上下。就以數學為例，15 至 11 級分的人數比例，由表 1，110 學年各佔約 1.26%、1.57%、2.51%、3.41%，及 4.57%；至於 109 學年，由表 2，則各佔約 11.22%、8.78%、6.84%、6.89%，及 7.05%。各級分的考生人數比例，在不同年度間，相差很大。甚至 110 學年的前 4 級分共佔約 8.75%，尚不到 109 學年 15 級分所佔約 11.22%。至於 5 標之級分，由表 3，110 學年分別為，11、9、6、4、3；由表 4，109 學年則為 14、13、9、5、4。110 學年較 109 學年，頂標降了 3 級分，前標降了 4 級分，均標降了 3 級分。兩年內 5 標的級分落差這麼大，級分及 5 標怎能說是穩定呢？另外，對於學測，若有人問 8 級分算好還是不好？被問者可能將瞠目結舌，不知該如何回答。以 110 學年為例，由表 3，8 級分在國文為底標，社會為後標，都不能說太好；但英文及自然皆為均標，8 級分算是普通；至於數學則屬不錯，因前標為 9 級分、均標為 6 級分，8 級分可說接近前標了。所以不同科目的同一級分，即使在同一年度，也無法相對照。

前面指出，因科目或年度的不同，級分並無法相比。但若各科級分，是將原始成績排序後，依到考人數等分，則有如百分位數的原理，即使不同科目或不同年度，同一級分所反映考生的表現優劣，便較接近。不時會看到媒體報導，某

項考試成績的分佈為單峰，或雙峰(又稱 M 型)。不論單峰、雙峰，或其他，由此不難得知，成績分佈常是不均勻的。另外，為讓人了解國民的薪資高低，除中位數外，行政院主計總處亦會提供幾個百分位數，如第 10、第 20、⋯，第 90 等，以讓人知道薪資高到那裡去，及低到那裡去。用百分位數，而不是從最低者(或最高者)，以一定額來升一級(或降一級)，如年薪 50 萬、60 萬、70 萬，⋯，道理很容易想通。因而目前這種以“分數等距”來切割級分，並不易了解如此轉換的邏輯。級分是按分數切割，5 標則是按人數切割。忽而按分數忽而按人數，要知各科 5 標雖依相同的百分位數，但由於已先經級分扭曲原始成績，對每一級分，其分量(指所佔考生人數之百分比)差很多，造成不同科目或不同年度的同一標，並無法相對照。我們猜想，除了瞎掰外，將成績這樣的先級分再 5 標的轉換，不會真有什麼科學依據。那為什麼硬要創造出級分及 5 標呢？

沉且，我們早已提醒，在“9 年課綱”的“細目銓釋”裡，於“能認識百分位數的概念”項下，指出“知道百分位數通常用於分析總次數多的資料，避免在資料數少的例子中，做百分位數的教學。”其中“知道”應是指要讓學生知道。如果連學生都該知道，“大考中心”的眾專家豈能不知？從已過時的“9 年課綱”起，便強調數據少時，就不適合採百分位數。我們還曾舉例說明，當數據較少時，便會有很多百分位數取同一值。也就是當數據不多時，若求百分位數，會弄得彆彆扭扭的。有人可能反駁，學測考生那麼多，

數據怎會少？學測報名人數，在 109 學年，有 13 萬餘，110 學年，則有 12 萬餘，的確不少。但不要忘了，不要說各科原始滿分不過 100、128，或 144，本就一點都不多，將成績轉換成級分後，連同 0 級分，相當於一筆資料數只有區區 16 的數據。資料數才 16，對其用百分位數分割，5 標中若有兩標是同一級分，一點都不必太驚訝，那絕非稀有事件。

我們說過，頂標、前標、均標、後標，及底標等 5 標，分別依到考生之第 88、75、50、25、12 等百分位數之級分而定。110 學年學測數學各級分之百分比及累計百分比如表 1。首先因四捨五入之故，表 1 裡各級分百分比之和為 99.99，但這並不會造成什麼影響。因底標為第 12 百分位數，而 0 至 3 級分累計 17.83%(≥12%)，至少 3 級分累計 92.30%(≥88%)，於是底標為 3 級分。因後標為第 25 百分位數，而 0 至 4 級分累計 30.24%(≥25%)，且至少 4 級分累計 82.17%(≥75%)，故後標為 4 級分。因均標為第 50 百分位數，而 0 至 6 級分累計 55.16%(≥50%)，且至少 6 級分累計 56.72%(≥50%)，故均標為 6 級分。因前標為第 75 百分位數，而 0 至 9 級分累計 80.96%(≥75%)，且至少 9 級分累計 26.09%(≥25%)，故前標為 9 級分。因頂標為第 88 百分位數，而 0 至 11 級分累計 91.25%(≥88%)，且至少 11 級分累計 13.33%(≥12%)，故頂標為 11 級分。原本設定 5 標各為第 88、75、50、25、12 百分位數，因而達到頂標至底標的各標人數百分比，分別預設為 12%、25%、50%、75%、88%，如今由表 1，110 學年，數學實際的 5 標人數百分比，由頂標至底標，

各為 13.33%、26.09%、56.72%、82.17%、92.30%，與預設的百分比，差異算是不太大。

國文呢？表 5 給出 109 學年，學測國文各級分之百分比及累計百分比。可求出底標為 8 級分，後標為 9 級分，均標為 11 級分，前標呢？由於 0 至 12 級分累計 73.43%，差 75% 一些，只好再加一級分，而 13 級分有 14.89%，至 13 級分累計後，成為 88.31%，不但超過 75%，也超過 88%，遂造成前標及頂標皆為 13 級分。這可非偶然，108 學年學測國文的前標及頂標也皆為 13 級分。再度，這些數字，由於有四捨五入之故，可能會有點小誤差。

由表 1、表 2 及表 5 可看出，各級分人數所佔百分比，差異相當大。某一級分人數所佔百分比過大，便有可能造成 2 標同一級分。附帶一提，由某科的到考人數，如 128,424 人，

$$128{,}424 \times 0.88 = 113{,}013.12 \rightarrow 113{,}014。$$

從最低往上數之第 113,014 位到考生的級分，便為頂標，餘類推。不過因有如表 1 之各科級分累計表，由此便可求出 5 標，這樣簡便多了。

表 1. 110 學年學測數學各級分之百分比及累計百分比

級分	百分比	累計百分比
15	1.26	100.00
14	1.57	98.74
13	2.51	97.17
12	3.41	94.66
11	4.57	91.25
10	5.71	86.67
9	7.05	80.96
8	8.58	73.91
7	10.16	65.32
6	11.88	55.16
5	13.05	43.28
4	12.40	30.24
3	10.14	17.83
2	6.24	7.70
1	1.42	1.46
0	0.04	0.04

表 2. 109 學年學測數學各級分之百分比及累計百分比

級分	百分比	累計百分比
15	11.22	100.00
14	8.78	88.78
13	6.84	80.00
12	6.89	73.16
11	7.05	66.27
10	6.24	59.22
9	7.28	52.99
8	7.11	45.70
7	6.05	38.60
6	7.05	32.55
5	7.06	25.50
4	6.45	18.43
3	6.92	11.98
2	4.13	5.07
1	0.89	0.93
0	0.04	0.04

表 3. 110 學年學測各科成績標準

標準 項目	頂標	前標	均標	後標	底標
國文	13	12	11	9	8
英文	13	12	8	5	4
數學	11	9	6	4	3
社會	13	12	10	8	7
自然	13	12	8	6	5

表 4. 109 學年學測各科成績標準

標準 項目	頂標	前標	均標	後標	底標
國文	13	13	11	9	8
英文	14	12	9	6	4
數學	14	13	9	5	4
社會	13	12	10	8	7
自然	13	11	8	6	5

表 5.109 學年學測國文各級分之百分比及累計百分比

級分	百分比	累計百分比
15	3.26	100.00
14	8.43	96.74
13	14.89	88.31
12	17.66	73.43
11	16.33	55.77
10	13.38	39.44
9	9.05	26.06
8	5.92	17.01
7	4.00	11.09
6	2.76	7.09
5	1.94	4.32
4	1.35	2.39
3	0.81	1.04
2	0.21	0.23
1	0.01	0.02
0	0.01	0.01

13 再談級分

“治大國若烹小鮮”，出自老子“道德經”。意思是說，治理大國與煎小魚一般，不能常去翻動，否則小魚將破碎而不成形了。2021 年 8 月 3 日，聯合報有則“明年分科測驗混合題配分出爐”之報導。一開始便寫著，“配合新課綱，大學指考明年更名‘分科測驗’新登場，各科都將新增雜揉選擇題與非選題的‘混合題型’。有高中教師形容考生將很忙，除畫答案卡，還要簡短描述自身觀點、製作表格、在地圖上作圖等。”‘混合題型’讓考生很忙，恐怕還不算太大的改變。

國內大學招生，從 1954 年開始有大學聯考，1994 年導入多元入學，也開始有學測，如今自 2022 年起，入學方式將有更大的改變。本來只有學測才將考生原始成績換成級分，每年 7 月的指考(指定科目考試)則仍以原始成績呈現。安定了幾年，“大學招聯會”(大學招生委員會聯合會，由參與大學招生之各大學校長組成)可能覺得不能讓師生太安逸，該“行拂亂其所為，所以動心忍性，增益其所不能”。於是自 111 學年度起，考招將有新制，不但將大學指考更名為“分科測驗”，共考數學甲、物理、化學、生物、歷史、地理、公民與社會等 7 科，且每科滿分將由 100 分改為 45 級分。另外，學測也將由現今的 15 級分，改成 45 級分。此大變革早於 2017 年 3 月，經“大學招聯會”通過，緩衝幾

年，自 2022 年起實施。

科舉時代，考試沒有英文、數學、社會及自然，國文呢？不完全算有，因就只考個作文。歷來在不同的考制、不同的價值觀下，出頭的人便大不相同。時不利兮騅不逝，生不逢時也沒辦法。就如愛國詩人杜甫(712–770)，在他的時代，他雖 7 歲能作詩，但“舉進士不中第，困長安”，並不得志。若生在今日呢？他那些“國破山河在，城春草木深，感時花濺淚，恨別鳥驚心”之類的詩，恐怕會被某些人痛罵這麼嫌棄本國就離開吧！豈可能被奉為詩聖？

眼看 2021 年 9 月升高三的學生，即將適用新制，不少家長及老師不願認命，開始表達關切。依媒體報導，贊成級分制，及改成 45 級分者，有認為“可避免考生因些微分數差距，失去適性發展的機會。”有認為“可避免因各科性質不同、題目難度不同，對部分考生產生不利。”也有認為“採用級分制才是真公平。若將不同科目的原始總分(百分制)相加，就像把賽跑選手在 100 公尺、400 公尺競賽的秒數直接加在一起，忽略了每一科的每一分，代表的意義不同。”反對者有認為“計分粗略，恐導致同分者眾，不利選才。”有認為“若執意分科測驗改為 45 級分新制，恐將產生不公平，包括分數倒置造成不公，亦恐出現超額問題。”亦有認為“級分制以各科前 1%考生的原始成績作為級距，各級分累計人數會隨當年考試難易度而變動，且都是整數，無法計算到小數點以下。這樣的評分標準不穩定，區分考生優劣的方法過於粗略，將帶來更多不公平。”

贊成改為級分制者，所提理由都是不堪一擊的，如說“將不同科目的原始總分(百分制)相加，…，忽略了每一科的每一分，代表的意義不同。”只是連行之多年的學測，都做不到每一科的每一分，代表的意義相同，“分科測驗”怎就做得到？我們已指出，學測在不同年、不同科的每一級分，代表的意義差異可能極大；且學測命題難易程度向來並不穩定，怎會因轉換成級分後，每一分代表的意義便相同？2020 年，“大考中心”主任，還因數學科滿級分人數超過 1 萬 4 千人，遭抨擊缺乏鑑別度，而請辭獲准。

由於反對聲音不小，經教育部邀各團體代表開會後，“大學招聯會”讓步，同意進行研議 2022 年，各科採 45 級分、60 級分和 75 級分的可能性，不排除修改既有規畫。不知“大學招聯會”何以對級分那麼執著？當初學測會訂出各科皆採 15 級分，應是有經多方討論。2017 年“大學招聯會”，通過將大學指考每科改成 45 級分，想必亦該經深思熟慮，且該有學理依據。如今實施在即，遇有人反對，便願意考慮改採 60 或 75 級分，那豈不顯示當初的思慮不周嗎？

“大學招聯會”的考慮結果如何？對“大學招聯會”而言，朝令夕改還真的不困難，3 個多星期後，於 8 月 26 日宣布，將調整為每科 60 級分，也就是 2017 年所訂每科 45 級分的改變被推翻了。“大學招聯會”說明其依據是，“將成績轉換為 45、60、75 三種級分模擬分發，最後確認 60 級分對降低同分增額情況更有效”。奇怪，當初未做此模擬嗎？而且，天下本無事，維持每科原始成績，對降低同分增

額的情況，不是更有效嗎？很多家長團體對“大學招聯會”此由 45 至 60 級分的讓步，自然仍不滿意，主張維持現行每科 100 分的“百分制”。

現況當然可以推翻，在統計學裡，一切都是假設，就看接受那一個。當懷疑現況時，可做一假設檢定(testing hypotheses)。在假設檢定裡，秉持無罪推定的精神，盡全力維護現況，只有在證據相當顯著時，才推翻現況，接受新假設。朝令夕改絕非統計的精神。有人可能質疑，朝令有錯夕即改，善莫大焉，不是該被讚美嗎？非也！由朝到夕，在這麼短的時間內便能發現有錯，那很可能是當初的“令”思慮不夠周詳。既然過去有思慮不周詳之嫌，當然不被輕易相信能很迅速便英明起來。這種保護現況的精神，是做決策者，都該奉行的圭臬。惟有一旦通過的方案，便很難更動，做決策時才會儘量謹慎。負責考招的單位，看來宜“加強假設檢定裡的統計思維”。

在“論語”“子張”篇裡，“子夏曰，‘雖小道，必有可觀者焉；致遠恐泥，是以君子不為也。’”意思是說，即使是小招數，也必有可取處。但若想以之來達到令人刮目相看的績效，將有如雙腳沾滿泥土，走不遠的。所以君子才不樂於花心思在那些小眉小眼的事務上，以免陷入泥沼，一無所成。負責為國取才者，如果只習於在級分這類小玩藝上打轉，那真是會致遠恐泥了。

從拿到什麼試題、臨場反應，到志願的選擇，本來考試

就免不了有隨機性，那算是無可奈何，考生不得不承擔。現今學測將原始成績轉換成級分，相當於考生又得多承擔一隨機性。但這是“大考中心”強加的。只是“大考中心”為何要將考生的原始成績，做此擾動呢？

由於“大考中心”依分數等距，而非依人數百分比，將原始分數轉換成級分，造成考生成績排名可能超乎想像地大翻轉。假設有 A、B 二生，皆參加某學年之學測。國文、英文、數學、社會及自然等 5 科，A 生比 B 生的原始成績，有 4 科各少 1 分，有 1 科多 8 分，因而原始總分，A 生比 B 生多 4 分。若看原始成績，不會認為 A 生遜於 B 生，甚至可能以為 A 生優於 B 生。經“大考中心”換成級分後，A 生 5 科中，有 4 科 14 級分，1 科 15 級分，總級分 71；B 生則 5 科均為 15 級分，總級分 75。因 B 生較優的 4 科，均在 15 級分邊緣，因而 A 生雖各科只差 B 生 1 分，便都落入 14 級分了。至於 A 生有一科多 8 分卻沒啥大用，因與 B 生同為 15 級分。不是為了讓學生不要太計較分數，才採級分制嗎？如今總分多 4 分，卻反而少了 4 級分，能選的校系差很多，怎能不計較呢？這種例子很多，主因是原始成績只要差 0.01 分，便可能差 1 級分，而總有這種時運不濟者。事實上，原始總分較高，總級分卻反而低了 4 級分，看起來很傷，卻還不是最離譜的。各位不妨檢視歷年來的學測成績統計，原始總分較高，總級分要低到 8 級，都有可能。例如，有某生 4 科原始分數極高，幾乎滿分，因而都 15 級分，但 1 科考壞了，只有 7 級分。該生原始總分仍相當高，比很多 75 級分

者高，但總級分卻只有 67，差了好幾個檔次。每年有超出10 萬個考生，有幾位運氣如此不佳者，差不多可說是必然。

原始成績何以要全換算成 15 級分？“大考中心”的說詞是，“為了不要讓考生為了 0.1 分的差距而計較，於是捨棄傳統以原始分數為直接成績標準的方式，而改用級分制”，但此理由並不易被接受。“大考中心”如此翻轉成績排序，究竟有什麼目的？又能達到什麼效果？難道是主導學測命題的“大考中心”，覺得各科試題之鑑別率不高，遂以級分來增加一隨機性，以降低試題之影響？

早期考完試，若分數普遍太低，教師有時會將成績調整，以開根號乘以 10。即將原本的 x 分，換成 $10x^{1/2}$ 分。採用此法的主因是，簡單、分數仍介於 0 到 100 間，且高低分順序不變。缺點是，每人加的分數不同。考 25 分者加最多，變成 50 分了。但 0 分仍是 0 分，100 分仍是 100 分。而本來考 1 分與 0 分，差別極小，但經換算後，前者成為 10 分，後者仍是 0 分。這樣加分有沒有道理？沒有！只不過讓老師方便加分。又，雖每一科的高低順序維持不變，但若幾個老師都採此成績調整的方式，則總分便有差了。例如，假設有 D、E 二生，考國英數 3 科。D 生分別得 81、64 及 36 分，總分 181；E 生分別得 100、100 及 4 分，總分 204。原本 E 生總分較 D 生高出 23 分，遙遙領先。但 3 科都經開根號乘以 10 的調整後，D 生成績分別為 90、80 及 60 分，總分 230；E 生則分別為 100、100 及 20 分，總分 220，反而落後 D 生 10 分。

對於學測，如果是將原始成績排序後，依到考人數等分來給級分，還可想出理由。但“大考中心”至目前，均以將原始成績依等距來畫分級分，此乃是在做一件目的不明之小道。尤其即將做的指考成績之轉換，若連 75 級分也列入考慮，那 100 級分難道不行嗎？而若可以有 100 級分，原始成績又為什麼不行？

級分是高中生申請入學用的。高中生可提供中學時代的各種表現，包括各項競賽成績。大學校系在辦理甄選時，可依自己學系的屬性，自行評斷申請學生提供的每份資料之價值。何以重要的學測成績，這是每位申請者皆有的，齊一標準的表現，卻被棄若敝屣，只能提供被加工過後的成績，也就是級分？考生種種差異極大的“優異”表現，各學系能自行判斷其分量，學測的原始成績為何便不行？“大考中心”只要提供各科成績累計的百分比，供各學系參考不就好了。此正如若要比較芒果與鳳梨何者較營養，應提供新鮮的芒果與鳳梨，而非加工後的芒果乾與鳳梨乾。而指考行之多年的各科滿分 100 分，也想改成級分制。“大學招聯會”說，“無論一科幾級分，維持‘級分制’是他們堅守的最後防線，級分制才能落實科科等值，不讓考招走回現行‘百分制’帶來的分分計較情況。”科科等值？不分分計較？“大學招聯會”都不看過去學測‘級分制’分分計較的數據嗎？負責為國取才者，變成級分控了。

14 門檻與篩選倍率

1968 年，台灣開始有 9 年國教，初中聯考至此結束，之後只有高中聯考及大學聯考，當然另還有專科等體系的聯考。早期除極少數的保送名額外，要進入高中或大學就讀，除參加聯考外，沒有其他管道。那時錄取率低，學子拼命苦讀，三更燈火五更雞，參加筆試後放榜，幾家歡樂幾家愁。落榜或認為考得不滿意者，只好明年捲土重來，再拼一次，考到第三次的也不算罕見。行之多年後，聯考被批評乃一試定終身，罪大惡極。於是多元入學方案上場。既是多元，負責單位的認知，似乎便可以一再修改入學方式及擴展入學管道。朝三暮四及朝四暮三，便能將猴子搞糊塗，時至今日，各種多元入學方案，已到令人眼花撩亂的地步，並持續更撩亂中。處於學校多學生少，在錄取率一年比一年高的當下，進高中及大學，早非難事。但學生心目中，適合自己就讀的學校，標準也逐漸提高了，因而競爭依舊在。另一方面，雖有多種方式入學，並不表學生會變輕鬆。因不過是將一塊餅，第一輪時先切下一部分，讓學生爭奪，再來第二輪，說不定還有第三輪。每次爭奪的方式不同，若爭不到第一塊，或對爭到的那片不滿意，學生只好迎接下一戰。但第二戰的爭奪方式乃大不相同，有時還會修改規則。因而學生不但比過去更辛苦，還要隨時留意作戰方式是否有變。

這其中常有個篩選，先來看個例子。剛才提到，本來欲

進高中，只有參加聯考一途。二十多年前，在國三下開始有推甄。台灣各縣市的各國中，依分配到的名額，各自推薦校內國三學生，參加該縣市高中的甄試。推薦方式大抵是依學生在校 5 學期的成績而定。在某大城，男女各推薦出約 1 千名學生後，先參加欲進高中之筆試，考國文、英文、數學、社會及自然，各科滿分皆為 1 百。某著名女中，對社會及自然設下門檻，兩科皆要達到所有推薦生的 33 百分位數以上，然後採計國文、英文及數學 3 科的總分。全部約 1 千名參加推甄的女生，約 3 分之 1 可被錄取該著名女中。放榜後有些原本相當有把握的學生落榜了，這令其頗為訝異，待收到成績單後則感到傷心。原來她們國英數的總分，有比錄取標準多二十餘分者，但自然比門檻少了些，甚至有只少 0.2 分者。該校校長說明，女生有不少不太喜愛自然，社會則有些學生以為不太重要，該校重視國英數，但希望學生不要放棄社會及自然，因而採這種甄選方式。

雖該女中依其理念訂出甄選程序，只是很難說那是合理的。因自全縣市 1 萬多名國三女生中，推薦出才約 1 千名學生(約佔全縣市國三女生的 8%)，她們第 33 百分位數成績的意義，與學測十幾萬考生的 5 標之意義完全不同。若所有全縣市的國三女生都能應考該女中的甄選，則該女中原本所訂的自然與社會之門檻(全縣市女生約前 8%的前 2/3，乃高於全縣市女生的前 6%，即約第 94 百分位數)，將比全縣市女生的頂標(第 88 百分位數)高多了。該女中對自然與社會設這麼高的門檻，已非僅是要求學生勿放棄此二科之原意。另一方

面，該女中較重視的國英數 3 科，由於是看總分，因此若有考生國英數中有某科，比參加甄選的那約 1 千名學生，其第 33 百分位數的成績少 20 分也沒關係，因只要 3 科總分足夠即能被錄取。但總分不採計的社會與自然，卻只要比第 33 百分位數的成績低，便被刷掉了。這樣的思維，顯然有違該女中所宣稱較重視國英數的本意。其實，若對某科訂個不低的門檻，便表示該女中並非不重視該科。而既然對某些科目訂下不低之門檻，存在上述缺點，那便不必有什麼門檻了，逕自對所有該校“在乎”的科目，於總分採不同的加權(即加重計分)，而該校以為“不在乎”的科目，則不列入總分之採計。這應才較符合原本對某些科目訂門檻，卻不列入總分採計的該校之思維。

只是這種不盡合理的訂門檻方式，並非只在擬進高中時才有，直到今日，大學校系也普遍存在。事實上，學測的 5 標，就是“大考中心”讓各學系於辦理個人申請入學時，依據來訂檢定標準的。不能用級分訂，只能用 5 標，“大考中心”可說管很多。只是由於各標的標準並不穩定，如 106 至 110 學年，數學的頂標級分，分別為 11、12、14、14 及 11，而所佔人數之百分比，則分別為 14.38%、14.88%、12.44%、20.00%，及 13.33%，符合各標考生的級分及人數百分比，各年間有不小的差異，也就是每年檢定標準的門檻寬鬆不一。不但如此，偶而還會有某科兩標同一級分的情況發生。即有些學系精心討論，以決定對某科該訂頂標或前標，因各有優缺點，事後卻可能發現，花長時間討論根本是多餘，因頂標

及前標竟然同一級分。

每一大學校系的報考生，在第一階段通過檢定標準後，尚須被篩選出來才行。各學系在訂定各科篩選倍率時，不知是否了解“大考中心”的篩選順序，是從倍率高的開始？而這會造成什麼影響？在此以某大學理學院某學系為例。該系申請入學的招生名額為 22，而第一階段檢定標準是國文、英文、數學，及自然皆均標，篩選倍率則國文 10 倍、英文 8 倍、數學 6 倍，及自然 3 倍。“大考中心”的作業程序如下：因招生名額 22，先從通過檢定標準的申請者中，國文最高的，挑出 220(=10×22)位；再由此 220 位中，從英文最高的，挑出 176(=8×22)位；然後依數學，從中挑出 132(=6×22)位；最後依自然，從中挑出 66(=3×22)位，以進入第二階段。也就是說，先挑國文較好的；再從中挑英文較好的；再從中挑數學較好的；最後才比該系最重視的自然。只是總會有考生，在比不到自然之前，便早被剔除了。是國文不夠好的勿來，而非自然好的國文不能差。這樣的篩選程序，不見得是該系的選才思維。申請者的國文、英文、數學，及自然等 4 科，於通過檢定標準後，還得通過該系依序所設的國文、英文、數學，及自然等 4 道篩選關卡。以 104 學年度為例。此 4 科之均標級分，依序為 11、9、7 及 9，不算太高。至於 4 科通過篩選的最低級分，依序為 12、13、13，及 14，不但皆高於檢定標準，且英文、數學及自然，還高出不少。由此可看出，通過該學系所設的檢定標準，根本沒啥用。而如果有位考生，此 4 科的級分分別為 11、15、15，及 15，便通不過篩

選關卡，因第一科國文 11 級分，雖已達該系的檢定標準 11 級分，卻比不過該系其他申請者的 12 級分(這時級分便可派上場了，但檢定標準只能採 5 標之一)，因此在第一階段早早就被刷掉了。不要說英文、數學，及自然都不必比了，就因國文的差 1 級分，便沒機會參加真正比賽的第二階段之甄選。要通過申請以進入該學系，最關鍵的科目，居然是該學系不以為須太重視的一科國文。

對上述學系，再以 110 學年度為例。此 4 科之均標級分，依序為 11、8、6 及 8，至於 4 科過篩選的最低級分，則依序為 11、10、10 及 14。如果有位考生，此 4 科的級分分別為 11、9、15 及 15，便通不過篩選，因第二科英文 9 級分，雖已達該系的檢定標準 8 級分，卻比不過該系其他申請者的 10 級分，因此在第一階段就已被刷掉了。而即使過了 4 科門檻及篩選倍率仍早得很，因這僅是第一階段。一旦進入該系第二階段的甄選，還要加權 4 科學測級分：國文×1+英文×1.5+數學×1.5+自然×1.5。你看！先各科門檻，再各科依不同的倍率篩選，然後再比加權總分，把資訊已縮減不少的學測級分成績，一關又一關，拼命想從中榨出些什麼資訊。但其實加權總分並沒那麼關鍵，因不過佔第二階段總成績的 40%。另外尚有審查資料及口試(104 學年兩者各佔 30%；110 學年前者佔 40%，後者佔 20%)。而在這冗長、反覆又複雜的過程中，卻只要在該系認為份量不高的國文或英文少個 1 級分(而這可能是因國文或英文的原始成績少 0.01 分)，便全都沒有後續了。

對前述學系，篩選倍率最低，理應是該系愈在乎的科目，因而該學系最在乎的科目，依序應是自然、數學、英文，而國文當然便算普通在乎了。只是若在 104 學年自然、數學及英文皆滿級分的考生，而國文 11 級分，該系卻沒機會看一眼。該生擅長且該系認為較重要的自然、數學及英文等 3 科，完全無用武之地。被當做第一道關卡者，是篩選倍率最高的，是該系並不以為太要緊的一科，後果卻可能過慮掉學系應會相當欣賞的學生，這說不定是不少大學校系沒想清楚的。

相信有相當多的學系，在填寫篩選倍率時，並不太明白所填那些倍率之後果。簡單講，在國文較好的申請者中，挑選自然較好的，是否真為當初填寫篩選倍率時之本意？還有一點也值得留意。若填寫篩選倍率的只有 A 與 B 兩科，且 A 科為 4 倍，B 科為 3 倍，則“大考中心”先挑 A 科較高的 4 倍學生，再從中挑出 B 科較佳的 3 倍學生；但若 A、B 兩科之篩選倍率皆為 3，則乃依 A 與 B 的級分和來篩選。各學系在填寫倍率時，對於前者，想法可能是 B 科稍微重要些，結果卻幾乎是由 A 科便決定能進入第二階段的人選了；至於對後者，其想法可能是 A、B 兩科差不多重要，卻不見得認為兩科可互通，只需看兩科級分和。篩選倍率之些微差異，便導致篩選方式有極大的差異，這可能亦是不少學系始料未及的。另一方面，若只有數學及總級分有篩選倍率，皆設定為 3，則便依二級分和來挑考生。那便相當於在總級分裡，數學加重 1 倍，然後挑 3 倍的學生。但該學系真有這個意思嗎？

總之，各學系於決定篩選倍率前，對“大考中心”的操作方式，宜有充分了解，以確保所訂的倍率，果真是想要的甄選程序。

要知跑步固然是各項運動競技很重要的能力，要參加國際比賽英文好當然也有不少便利。但如果挑選舉重選手，第一關是先比 5,000 公尺，挑出最強的 20 名，於第二關比英文，挑出最強的 10 名，能過此二關者，第三關才比舉重，則所錄取者，是都有若干能力，卻不見得是最適合的舉重人選。

15 取樣

在“孟子”的“萬章篇上”有，“‘泰誓’曰‘天視自我民視，天聽自我民聽’”，其中“泰誓”是“尚書”之一篇名。即使上天知道世上苦人多，但總得先了解人間需求，才能造福或拯救世人，上天因而需要民視及民聽，以期至少達到所謂“下民易虐，上天難欺”之境界。只是從人的角度，上天的視與聽，可能屢屢不夠周全，或有所偏頗，因而怨恨老天無眼，或老天不公的世人，自古以來，從未少過。既然連老天都可能無眼，則凡人無眼，或者說不善取樣，便不足為奇了。

在“大不列顛兩千年：從羅馬行省、日不落帝國到英國脫歐，王冠下的權力更迭及對世界秩序的掌控”(The Story of Britain: From the Romans to the Present，2019，羅伊史壯(Roy Strong，1935−)著，陳建元(2021)為中譯本)裡，提到 17 世紀末，英國的政府逐漸龐大，卻能擁有高效率的體制。因他們“開始以反映當時科學進步的方式來讓政府運作，例如統計數據便在此時首次被使用”。這是指在英國，事實上，人類利用統計的歷史，乃相當悠久。

在“舊約聖經”第 4 卷“民數記”(Numbers)裡，第 1 章的標題便為“統計人口”(The Census)。摩西(Moses，約西元前 1520−1400 年)受上帝之命，率領被奴役的以色列人，離開埃及。歷經 40 多年的艱難跋涉，有時尚得仰賴神蹟。如當

前有紅海阻撓，後有法老王追兵時，摩西“舉手向海伸杖，把水分開”。憑此神蹟，以色列人遂得以下海中走乾地，終於到達富饒的迦南(Canaan，指地中海東岸的沿海低地，約今日的以色列一帶)。

安頓後，隔年耶和華曉諭摩西，要他點數各宗族 20 歲以上，且能服役的男丁。12 個宗族中，流便(Reuben)子孫的後代，共有 46,500 人；西緬(Simeon)子孫的後代，共有 59,300 人；…；拿弗他利(Naphtali)子孫的後代共有 53,400 人。總共有 603,550 人。計數人口能到十位數，應算是相當精準了。摩西帶領以色列人離開埃及的時間，歷來有幾個觀點，包含西元前 1513 年、西元前 1445 年，亦有認為西元前 1290 年者。不論那個年份，均比中國的周朝(約西元前 1100−256 年)還早。因此這應是歷史上相當早期的一個人口統計，看起來乃採普查的方式，且應花了段不短的時間。因在“撒母耳記下”(2 Samuel)的第 24 章，著名的大衛王(King David，西元前 1040−970 年)在位期間(西元前 1010−970 年)也曾點數過人口，得到能作戰的男丁，以色列(Isreal)有 80 萬人，猶大(Judah)有 50 萬人。誰去點數的？大衛王指派他的親信元帥(army commanders)約押(Joab)，走遍以色列眾支派去執行。而就算有大元帥出馬，此人口統計，也花了長達 9 個月又 20 天！

為了賦稅、徵兵，或徭役之類的原因，自古以來，人口統計對治國是很重要的。早期是採普查的方式，且得派個幹練又有實力的親信去執行。而總共才 130 萬人，又是帝王時代，也要歷經約 290 天才能點數完，顯示人口普查之不易。

調查時間那麼長，不論再如何仔細，期間總會有死亡或遷徙，因此誤差在所難免。

時至今日，類如摩西及大衛王時的人口普查，已不常見了。取而代之的，是各式各樣的抽樣調查。以較好的方式，其準確度可能超過費時費力的普查。以座頭鯨(humpback whale)為例，這是鯨類中最巨大者。由於濫捕，數量逐漸降低，瀕臨絕種。近年來因保育成功，數量已慢慢回升。只是大海裡的座頭鯨究竟有多少頭呢？顯然無法以普查來計數。經由對攝影數據進行統計分析，可估計全球現存總數。

直觀上，取樣要客觀，所得結論才能較準確地反映現況。如政府推出各項重大政策前，利弊得失必有評估，但民意也要掌握。免得主政者被譏久居高位，早已不知民生疾苦，猶如現代晉惠帝(曾講出“何不食肉糜”的晉惠帝司馬衷(259−307)，可能是中國歷來皇帝裡，近年來最常被台灣政治人物提起的一位。但史家評晉惠帝其實只是“甚愚”，若要講昏君，歷史上比他昏庸百倍的皇帝，所在多有)。至於不是“甚愚”的掌權者，卻推出不合常理的政策，便可能被認為是因具私心，遂不顧大眾利益。2021 年，新冠肺炎疫情蔓延，重創了台灣的旅遊及餐飲等不少產業，欲振興經濟，究竟政府該推出的，是自己一再宣稱，可創造數千億元經濟效益的五倍券，或是如不少代議士所宣稱，民眾期盼的是以現金紓困？掌權者拿出的數據，相信的人多嗎？民意代表提供的資料，又有多真實？似乎呈現各說各話的局面。做決策前常要有數據，但數據從何而來？枕邊人、親友、心腹、民調，或

明查暗訪，都是可能的來源管道。而收集到的數據，能因而一窺全豹，或僅是以管窺天？能見微知著，或根本是以偏蓋全？要知數據就在那裡，但運用之妙，存乎一心。

先看 2021 年 8 月 6 日，中國時報有一則標題是“美不敢接種者 認為疫苗比病毒危險”(The unvaccinated still think Covid vaccines are a risk, survey finds)之報導。新冠肺炎全球大流行，美國有線電視新聞網(CNN)，於 7 月 15−27 日間，訪問了約 1,500 位美國成年人，其中有約 14%堅稱，“絕不接種疫苗”。這比例與 2020 年 12 月的調查結果接近，顯示即使過了 7 個月，對疫苗抱持堅拒態度者，仍不改初衷。尚未接種疫苗者，其中僅有 23%認為，疫苗對降低新冠肺炎的死亡率，非常有效或極為有效；但有 53%認為，疫苗對健康的威脅，遠比新冠病毒來得大。而已接種疫苗者，則有 88%認為，感染新冠肺炎對健康帶來的風險，遠大於接種疫苗。在尚未接種疫苗者中，有 57%認為，新聞誇大新冠病毒之影響；而已接種疫苗者中，有 53%認為，媒體報導如實，24%認為，媒體低估疫情嚴重性。

上述調查顯示，已接種疫苗者，與尚未接種疫苗者，這兩群人對疫苗及疫情的意見相當不同。說起來病毒之影響力，及疫苗之有效性，應屬科學能判定的範疇，但仍會有南轅北轍之看法。所以取樣是很重要的，如果取樣偏差，即使樣本數再大，仍可能導致以偏蓋全的結論。另外，現今常利用網路做民調，但會主動上網填卷者，可能是對此議題較有興趣者，且可能以熟悉網路使用者居多，這樣收集到的意

見，便得提防是否偏頗。但若視為一種快速且低成本的調查，並未太嚴肅地看待，也就無妨。

聲譽卓著的 BBC(British Broadcasting Corporation，英國廣播公司)，於 2002 年 11 月，對英國境內多達百萬名的聽眾與觀眾，票選有史以來最偉大的 1 百位英國人。結果曾任首相的邱吉爾(Winston Leonard Spencer Churchill，1874−1965)排名第一，囊括約 44.7%的票。科學家牛頓(Issac Newton，1643−1727)，則排名第六。隔年 8 月，BBC 又對英國境外重新票選，結果牛頓躍升至第一，獲得約 21.6%的票。邱吉爾則落居第二，得票率降至約 16%。雖樣本數不少，但同一家廣播公司的聽眾與觀眾，在英國境內與境外，看法便可能差異很大。民調裡往往會做交叉分析，正是為了比較不同族群(包含性別、年齡、教育程度，及政黨傾向等)間，意見是否有所不同。

我們再來看媒體上，兩則有關民調報導的最後總結。第一則為，“該份民調係由…，負責問卷設計、調查、統計抽樣分析。調查範圍為…，調查方法是以‘電話簿抽樣法’進行，以台灣地區(不含金門、馬祖)年滿 20 歲以上之成年人為調查對象；以該年度‘中華電信住宅部電話號碼簿’為母體清冊，依據各縣市電話簿，所刊電話數佔臺灣地區所刊電話總數比例，決定各縣市抽出之電話個數，為求完整的涵蓋性，再以系統抽樣法抽出各縣市電話樣本後，隨機修正最後 2 碼及 4 碼，以求接觸到未登錄電話的住宅戶。電話接通後，再由訪員按照戶中抽樣的原則，抽出應受訪的對象進行訪

問。為使樣本結構與母體結構更符合，本研究對樣本的分佈特性使用多變數反覆加權法(raking)進行加權。其中性別、年齡、教育程度及地理區域之權值，是依據內政部所出版‘中華民國臺閩地區人口統計’。”第二則為，“本次民調由…民調中心，負責問卷設計、調查執行、統計分析。調查期間於…進行，調查方法是以‘手機簡訊’進行，以台北市 18 歲以上民眾為調查範圍及對象;調查回收有效樣本數為 1,214 份，抽樣誤差在 95%信心水準下，為正負 2.81%。”

第一份民調，是從電話簿來抽樣，而為彌補未登錄電話者，對抽到的電話號碼，做了修正。不願登錄電話的族群，其意見說不定與登錄電話的族群差異很大，不可忽略他們。另外，電話接通後，也並非就以接話者為受訪對象(雖然這較方便)，而是有一套規則，免得若家中電話，通常由某位成員(如女主人)接聽，如此得到的樣本便有偏差。最後，還要為性別及年齡等因素來加權。凡此種種，都是想消除取樣偏差。即使如此，對家中未裝電話的族群，仍無法調查到。第二份民調，講得較簡略。與第一份民調大不相同的是，乃以手機來抽樣。但這便會漏掉未使用手機者。此則報導，有給出成功樣本數(超過 1,068)，並附上誤差(小於 3%)。事實上，電話能接通，且有人願接受訪問已不容易了，若還約束誰才是受訪對象，難免有不盡職的訪員，未遵守規定，而便宜行事。畢竟訪員有盡速完成成功樣本數之壓力。另外，由於為使成功樣本裡，各族群人數，符合母體裡，各族群之人數比例，執行民調的單位，有時會對收集到的樣本做加權。這本

有學理依據，但加權過程，卻常使執行者，被懷疑有可能趁機造假。

16 取樣偏差

“淮南子”乃西漢淮南王劉安(西元前 179−122 年)，召其幕下士人所完成，但掛名的只有劉安。如果在今日，劉安肯定會被批評違反學術倫理。但往昔並無著作權的概念，有人辛苦寫本書，惟恐自己人微言輕，心血將難被重視，還假借是某古代名人所作。人們常說的“一葉知秋”，便出自“淮南子”的“說山”篇，“嘗一臠肉，知一鑊之味；懸羽與炭，而知燥濕之氣；以小明大。見一葉落，而知歲之將暮；睹瓶中之冰，而知天下之寒：以近論遠。”意思是說，嘗一小塊肉，便知整鍋肉之味道；懸掛羽毛與木炭，便知空氣之濕度：這是由小明白大的道理。看到一片葉子凋落，就知道一年將盡；看見瓶中的水結冰，就知道天氣已很冷了：這是以近知遠的道理。

一葉知秋，與嘗一臠肉，便能知一鑊之味，原理相同。而二者之先決條件，皆是母體的變異不能太大。也就是若樹葉變黃脫落的時間，每年都差不多是進入秋天時，以及整鍋肉的味道均差不多，便容易一葉知秋，及一肉知味。生活上，一葉知秋的現象處處可見。我們有時僅由嚐一粒葡萄，便據此來判斷整串葡萄，甚至整批葡萄之酸甜，以決定購買與否。僅抽幾 cc 的血，便能對人體做檢驗。而對動物的捕捉−再捕捉法，通常也是基於某類動物，每隻被捕捉到的機率皆差不多。但對於人的各種調查，只要族群不同，差異性往往

不小，因此若取樣偏差，就不易以小明大、以近論遠了。

有些誤差是可容忍的。如大數法則指出，從母體隨機取樣(取出後放回，因而取出的樣本獨立且有共同分佈)，當樣本數夠大，則樣本平均接近母體平均的機率將很大。而我們曾說過，即使取出後不放回，雖樣本間並不獨立，但只要取出的樣本數，與母體數相比很小，則大數法則與中央極限定理，便仍可引用，誤差不至於太大。只是如果取樣過度偏差，則不要說將無法一葉知秋，有時甚至連落葉多至滿坑滿谷，都無法知秋了。因此在取樣時，不可不慎。

假設擬估計某大學男女生之比例，到校圖書館點數閱覽桌上的男女生人數，以所得之男女比來估計全校之男女比。這樣能一葉知秋嗎？不見得可以，除非已知男女生去圖書館的比例很接近。否則若有某一性別，較習於去圖書館讀書，則所得估計之誤差便將不小。另外，能以某大學圖書館借出書籍多寡之排名，作為該校學生喜愛看的書之排名嗎？並不盡然。因學生看書，除了向圖書館借閱，至少還有購買的管道。

在抽樣調查裡，樣本的產生必須很公正，才能獲得有效的推論。若在選樣過程中，有排除(或多取)某一類樣本的傾向，便稱“選擇偏差”(selection bias，即“取樣偏差”)。底下來看一著名的，於執行選舉民調時，因取樣偏差，造成預測大錯誤之例。1936 年的美國總統選舉，現任總統、民主黨籍的羅斯福(Franklin D. Roosevelt，1882−1945)競選連任，挑

戰他的是共和黨所推出，堪薩斯(Kansas)州的州長蘭登(Alfred Landon，1887−1987)。彼時美國正從經濟大蕭條(Great Depression)中復甦，雖全國失業率仍超過 10%，失業人口高達 900 萬，且人民的實際收入，比 1929−1933 年那段時期少了約 1/3，但情況正開始好轉。絕大部分的政治觀察家，均預測羅斯福將輕易連任成功，蘭登則將不堪一擊。不過“文學文摘”(Literary Digest)雜誌，卻力排眾議，於投票日(1936 年 11 月 3 日)前 3 天(10 月 31 日)，宣布他們的選前最後一次預測：蘭登將大勝羅斯福，普選得票率是 57%比 43%。另外，美國總統選舉，採特殊的“勝者全得”(Winner-take-all)制。各州各有一定名額的選舉人(elector)，所有選舉人，便組成選舉人團(Electoral College)。各州得票率最高的候選人，便囊括該州所有的選舉人票。那時在總共 531 張選舉人票中，“文學文摘”預測蘭登將獲得超過半數的 370 張，佔約 67.68%。也就是現任的羅斯福，將黯然神傷地下台。此預測雖讓人驚訝，大眾卻不敢視為企圖嘩眾取寵。因“文學文摘”並非等閒之輩，在美國總統選舉的預測，夙負盛名。此雜誌自 1916 年起，開始預測美國總統選舉，包括 1916、1920、1924、1928，及 1932 年，此前 5 次，從未失手。

即使過去的預測，有如神機妙算，“文學文摘”這回卻栽了個大跟斗，黯然神傷的是他們，而非羅斯福。因羅斯福獲得普選票的 60.8%，蘭登的得票率僅 36.5%。而除美國東北部新英格蘭(New England)地區的緬因(Maine)及佛蒙特(Vermont)兩個小州外，羅斯福在那時全美 48 個州中的 46 個

州勝出。選舉人票則獲壓倒性的 523 張，佔約 98.49%，蘭登僅獲區區 8 張，約僅佔 1.51%，離“文學文摘”預測的 370 張選舉人票，可說遠得很。不論普選票的 60.8%，或選舉人票的 98.49%，羅斯福的得票率，均是自 1820 年以來的最高。

“文學文摘”共寄出約 1,000 萬份問卷，回收約 227 萬份。那很可能是有史以來所做最大規模的一次民調，回收的問卷數也相當可觀。但怎會陰溝裡翻船呢？不是說只要成功樣本數有 1,068，抽樣誤便就 3%而已，如今成功樣本數有 1,068 的兩千多倍呢？不妨做個對照，前一年(1935 年)才成立的蓋洛普(Gallup)公司，那次僅用約 5 萬個樣本，便正確地預測羅斯福會贏，雖然得票率的估計有些誤差(預測羅斯福會獲得約 56%的票，實際得票率約 60.8%)。

“文學文摘”是如何挑選樣本的？從其訂戶、電話簿，及一些俱樂部的會員名單。他們的訂戶中，共和黨員的比例，較美國人口中之比例高，共和黨員當然傾向投給共和黨推出的候選人蘭登。又在 1936 年，美國電話尚非那麼普及(平均約每 4 戶才有一具)。此外，沒有參加任何俱樂部者，也被排除了。換句話說，抽樣過程中，有排除窮人的傾向：明顯的選擇偏差出現了。在 1936 年之前，這種選擇偏差，對於預測的影響還可容忍，因當時富人與窮人的投票行為，差異還不是太大。但在 1936 年，因經濟的因素，造成選民的政治傾向有很大的分野：窮人較多選擇羅斯福，而富人則較支持蘭登。當有選擇偏差時，樣本數再怎麼多，便可能都無用了。選擇偏差，是造成“文學文摘”那回預測，會謬以千里

之一主因。對“文學文摘”預測錯誤的檢討不少。選後 40 年，一篇於“美國統計學家”刊登的文章(Maurice C. Bryson (1976). The Literary Digest Poll: Making of a Statistical Myth. *The American Statistician*, 30(4), 184-185)指出，“文學文摘”該次誤差的實際原因(actual reason)為，其民調完全仰賴自願參與(voluntary responses)，因而回收的 227 萬份問卷，遠非隨機樣本，而是選民中，對該次選舉議題較感興趣者。而羅斯福的反對者比起其支持者，更積極參與民調。另外，再 12 年後，一篇於“公共輿論季刊”刊登的文章(Peverill Squire (1988). Why the 1936 Literary Digest Poll Failed. *The Public Opinion Quarterly*, 52(1), 125-133)指出，取樣偏差加上“不回答的偏差”(non-response bias)，造成“文學文摘”預測的錯誤。

我們略說明一下不回答的偏差。民調守則之一是，一旦決定受訪名單，就要儘可能去獲得他們的意見，一般這並不容易。當取出的樣本中，有過多沒有回覆或拒絕受訪者，將可能造成一嚴重的扭曲。有時不回答者與回答者的意見間，存在很大的差異。舉例而言，“文學文摘”在芝加哥(Chicago)所發出的問卷數約 240 萬，差不多是當時芝加哥全部選民人數的 1/3，因而絕不能說問卷數不多。但回覆者才約 20%！回收的問卷中，支持蘭登的超過半數。但選舉結果，羅斯福在芝加哥獲得約 2/3 之選票。另一方面，“文學文摘”對全國所發出的約 1 千萬份問卷中，只有比例不高的 22.7%回覆。這 22.7%回覆者的意見，不見得能代表全部被挑選出

來之 1 千萬選民。一般而言，低收入與高收入二族群，不回答問卷之比率較高。也就是回收問卷中，來自中收入者，往往超過該有之比率。而中收入者的意見，與高收入或低收入者，並不見得會相同。由於有這種不回答的偏差，現代民意調查機構，對於重大或較敏感的議題，在時間及經費允許下，傾向採用電訪(或面訪)，而非郵寄問卷。電訪(或面訪)之成功率通常能比郵寄問卷之回收率高很多。不過即使採用電訪(或面訪)，不回答的偏差仍然存在。那些電訪(面訪)時人不在家者，與在家者之習性，如工作類別、家庭狀況、社會背景等，可能會有很大的差異，想法因而將大不相同。拒絕受訪者的情況也類似。好的抽樣調查設計，會正視不回答偏差的問題，而採用較巧妙的方法，以設法降低不回答率。總之，抽樣調查設計，並不僅是統計上的問題，還要結合諸如心理學、社會學，及經濟學等方面的知識。

17 蓋洛普的總統民調

1935 年，年方 35，哥倫比亞大學(Columbia University)新聞學教授蓋洛普(George Horace Gallup，1901−1984)，創立蓋洛普(Gallup)公司。他利用統計學，以抽樣調查的方法，以進行民調，是美國科學民調的先驅。蓋洛普公司成立的隔年，1936 年，就憑僅訪問 5 萬位選民，即正確預測當年的總統選舉，羅斯福將擊敗蘭登，連任成功。初試啼聲便一鳴驚人，自此聲名大噪。美國從 1936 到 2012 年，歷經 20 次總統選舉，蓋洛普公司只有 3 次普選得票高低及當選者之預測均錯誤，分別是 1948、1976，及 2012 年。另外，2000 年之得票高低預測錯誤，但誰當選則預測正確。由於“勝者全得”選舉制度，美國總統選舉，候選人即使輸了全國普選票(popular vote)，仍有可能選上總統，歷來已有幾個例子。當選者正確預測 17 次，85%的成功率，算是不錯的。比較一下，若是瞎猜(表每次猜對誰當選之機率為 1/2)，則 20 次能至少對 17 次的機率為

$$(C(20,17) + C(20,18) + C(20,19) + C(20,20)) / 2^{20}$$

$$= 1{,}351/2^{20} \approx 1.288\times10^{-3}，$$

不到千分之 1.3 的機率，顯示蓋洛普能在 20 次當選者之預測中，正確 17 次，是有些真功夫的。另外，得票率高低正確預測率則為 80%，也仍不壞。但由此可看出，美國總統選舉

預測的不易。要知當兩黨候選人勢均力敵，一點風吹草動，便可能使豬羊變色。我們說過，民調就是有其侷限，所以預測不準，有時乃非戰之罪。百分之百正確的預測，反而可疑。底下分別來看蓋洛普對當選者預測錯誤之 3 次總統選舉。

1976 年，民主黨的卡特(James Earl Carter, Jr.，1924−)挑戰在位的共和黨福特(Gerald Rudolph Ford, Jr.，1913−2006)。蓋洛普預測普選得票率，福特會以 49.0%比 48.0%獲勝。結果卡特以 50.1%比 48.0%打敗福特。這是一場差距很小的選舉，預測有誤不好太苛責。

2012 年，蓋洛普預測共和黨的羅姆尼(Willard Mitt Romney，1947−)，會以 50.0%比 49.0%，險勝尋求連任民主黨的歐巴馬(Barack Obama，1961−)。但實際結果是歐巴馬以 51.1%比 47.2%順利過關。這次蓋洛普對得票率預測的誤差，比很多民調公司大。樹大招風，選後被有些專家，狠狠批評了一番。蓋洛普雖反駁其預測仍在統計誤差範圍內，但仍花了約 6 個月的時間自我檢討。之後，蓋洛普承認它的調查方法的確存在若干缺陷，包括在美國本土的東部與西部，撥打電訪的電話數都偏低、高估了白人的投票率、太依賴家用電話，使樣本裡年長者的比例偏高，以及在歐巴馬於“搖擺州”(Swing state)全力動員下，這些州的投票率顯著上升，而蓋洛普藉 7 個問題以判定選民是否會去投票，卻未能及時捕捉到這種變化。也就是之所以預測誤差過大，主要是對選民結構及投票行為之判斷有誤，加上取樣偏差。經此檢討，蓋洛普公司壯士斷腕，決定自 2016 年起，不再對美國總統選

舉進行預測，而致力於其他議題的民調。附帶一提，跟台灣各縣市類似，美國有些州長期支持民主黨，或共和黨，難以撼動。有些州則視候選人而可能變天，這些州便是所謂搖擺州。

1976 與 2012 年那兩次的預測不準，尚情有可原，因當雙方差距不大，估計便很容易失手。但 1948 年是怎麼回事？事實上，那次選舉，在競選過程中，現任總統民主黨的杜魯門(Harry S. Truman，1884−1972)一直不被看好。當時幾家主要的民調公司，包含蓋洛普在內，均預測挑戰者共和黨的杜威(Thomas Edmund Dewey，1902−1971)，普選得票率將贏杜魯門 5%以上。蓋洛普則預測杜威將以 49.5%比 44.5%獲勝(相加不到 100%，乃因另有其他兩位候選人)。結果是差不多反過來，杜魯門以 49.55%比 45.07%獲勝。選舉人票杜魯門得 303 張，杜威得 189 張，另一候選人得 39 張，合計 531 張。由於選前杜魯門極度被看衰，媒體對他遂百般奚落。有家報紙還曾以“DEWEY AS GOOD AS ELECTED, STATISTICS CONVINCE ROPER”(統計說服羅波相信杜威已當選)為新聞標題。羅波(Elmo Burns Roper, Jr.，1900−1971)是當時一位以預測精準著稱的民調專家。他預測杜威將以 52.2%對 37.1%，高達 15.1%之差距，擊潰杜魯門。事後看，當然覺得此預測錯得相當離譜，還拉統計背書，統計只能啞巴吃黃蓮。投票當天一大早，有家報紙已迫不及待地刊登一篇標題為“Harry S. Truman: A Study of a Failure”的文章。而投票還未截止時，有家信心滿滿的報社，已將隔天的頭版印好了，

標題是“DEWEY DEFEATS TRUMAN”。對現任總統如此羞辱，不知杜魯門有沒有懷恨在心，但至少沒聽說選後有那些媒體，遭當選的杜魯門控告散佈假消息。

1948 年美國那些民意調查機構，究竟為何犯下大錯？難道犯錯是因相信統計嗎？這未免令統計工作者難堪。原來當時從事民調，大都採“配額抽樣”(Quota Sampling，亦稱定額抽樣)。即欲進行調查，先將母體分類(層)，確定各類(層)的樣本數，在決定的配額數內，再經隨機抽樣，以產生樣本。舉一簡單的例子來看。某大學的大學部宿舍住宿生中，男生佔 60%，女生佔 40%，1、2、3、4 年級學生分別佔 40%、30%、20%、10%。今擬了解住宿生對餐廳之意見，依上述二類別，以定額抽樣方式抽取 500 個樣本，則男生樣本數為 300，其中 1、2、3、4 年級分別有 120、90、60、30 人；女生樣本數為 200 人，其中 1、2、3、4 年級分別有 80、60、40、20 人。

“配額抽樣”的原意，是想讓抽出的樣本，有如母體之一縮影，這本是一合理的想法。但在 1948 年的總統選舉，顯然是一失敗的經驗。怎麼會這樣？以蓋洛普為例，那時他們的民調採面訪方式。例如，一位在聖路易斯(St. Louis)城的訪員，要負責訪問 13 位選民，其中須包含：

(i) 6 位住郊區，7 位住城裡；
(ii) 7 位男性，6 位女性；

7 位男性中(對女性也有類似要求)：

(iii) 3 位在 40 歲以下，4 位超過 40 歲；

(iv) 1 位黑人，6 位白人。

6 位白人每月房租要滿足：

(v) 1 位超過 44 元，3 位超過 18 元未超過 44 元，2 位不超過 18 元。

在這麼多限制下，豈能再要求訪問之樣本須以隨機的方式產生？於是訪問名單，只要求各類配額皆達到即可，其餘便授權訪員自行決定了。蓋洛普希望樣本結構能忠實反映全部投票者的結構，所以才會顧慮各個主要會影響投票行為之特性。但其中卻有一罩門：無法對共和黨黨員及民主黨黨員給出配額。

要知在美國，只要在選前登記黨籍即可，初選時想投那一黨就登記那一黨，每次選舉可以不同。而一旦登記那一黨，就表支持該黨的候選人。只是政黨黨員比例，正是民調公司所不知道，而企圖藉由民調來獲知。精密給出各特性之配額，乃是想經調查後，能讓全國選民政治傾向的比例浮現。影響投票的因素其實很多，遠超過民調公司所能掌握的。由於訪問誰，本質上乃讓訪員自由挑選，但人的選擇，就是很難避免偏差。試想如果有學生的工讀，是在街頭發問卷，他難道不會傾向找看起來較順眼的人嗎？那些滿臉橫肉、殺氣騰騰者，任誰都避之惟恐不及，豈會趨前奉上問卷？另外，在調查時間壓力下，若遇一家人都願意填寫，難免會有訪員，不甩同質性過高的問題，在該戶多奉上幾份問卷。

在 1948 年，多數訪員挑出的面訪樣本裡，便有過多的共和黨黨員。一般而言，共和黨黨員較富有，且教育程度較高，因而電話擁有率亦較高(注意那是 1948 年)，又有較高比例的人，有固定的地址且住在較好的區域，導致較容易被訪問到。在此情況下，大部分的訪員，都可能在有意無意下，訪問了過高比例的共和黨黨員。

事實上從 1936 至 1944 年的總統選舉，蓋洛普對得票率之預測，都高估共和黨候選人，而低估民主黨候選人。也就是訪員有訪問過高比例共和黨黨員的傾向。只是在那段時期，民主黨推出的總統候選人，均為現任總統羅斯福，其實力遠高出共和黨的對手一大截。因此樣本中，共和黨黨員過多，所造成的偏差，尚能被掩蓋過去，因而使蓋洛普的預測誰當選依然正確。但在 1948 年，兩位候選人之支持度算是旗鼓相當，因此樣本的偏差，便影響到預測結果。

有了 1948 年的失敗經驗後，自此美國差不多所有的民調機構，皆以機率抽樣的方式產生樣本。只是美國幅員廣大，人口眾多(至 2021 年 9 月，美國人口估計已超過 3.3258 億)，採簡單隨機抽樣，難度極高。不要說不易拿到選民的底冊，就算有，抽出的選民分散在全美各地，從峽谷到高山，要進行訪問，不但困難且極耗成本。“叢聚抽樣”(cluster sampling，又稱部落抽樣，即將母體分成幾個群集(部落)，再從幾個群集中，抽出數個群集進行抽樣或普查。適用於各群集間的變異小，而群集內的變異大。當母體底冊之蒐集及編造相當困難，或想節省調查成本時，可採用此種抽樣)遂為一

常採取的作法。

直到 1980 年代中期，蓋洛普對總統選舉的民調，都採面訪方式，但逐漸困難起來。一方面成本愈來愈高；一方面在世風日下，基於治安的考量，民眾愈來愈不情願讓訪員進到屋裡去。但最主要還是面訪極耗時間及成本，當進入資訊在比誰能更快速取得的時代，面訪便完全不再具有優勢了。幸好電話日益普及，一般家庭都至少擁有一部電話。於是到了 1980 年代末期，蓋洛普的全國性民調，主要都採用電話訪問了。之後手機時代來臨，自 2008 年起，蓋洛普的民調方式，除電話外，已納入手機了。隨著通訊方式的多元，民調管道也將愈來愈多樣。又為了顧及老派人士，以手填寫的問卷，也仍然存在。

總之，民調要具參考價值，便得對母體結構更深入了解、樣本須更有代表性，且訪員的訓練要更紮實。而最不能忽視的一點是，對問卷的設計要更下功夫。

18 再談美國總統選舉

美國總統選舉，形式上乃採間接選舉，或者說委任直選。選民投票，其實是在選選舉人，再由選舉人依選民的意志，投票選出總統及副總統，有點迂迴。首任總統華盛頓(George Washington，1732−1799)，於 1789 年 4 月 30 日就職。自 1792 年起，每屆總統選舉的年份，都為 4 的倍數。自華盛頓於 1797 年 3 月 4 日卸任後，至 1933 年，新選出的總統，均為 3 月 4 日就職。其後自 1936 年起，新選出的總統，均於隔年 1 月 20 日就職，並延續至今。又，美國總統選舉，原本由各州自行決定投票日，1845 年，才制定全國統一於 11 月第一個星期一之後的第 1 個星期二投票(即如果 11 月 1 日為星期日或星期一，則投票日為星期二 11 月 3 日，或 11 月 2 日，其他情況則為 11 月 1 日後的下一個星期二)。如 2020 年是選舉年，而 11 月 1 日為星期日，故投票日為 11 月 3 日星期二) 。統一投票日，以免早投票的州對晚投票州的選民產生影響。只是美國幅員廣大，橫跨好幾個時區，就算同一天投票，當東部各州，已開始開票時，西部各州及夏威夷(Hawaii)州，仍在投票。

為什麼總統選舉投票日選在 11 月？在早期的農業社會，11 月時，繁忙的收穫季節已結束了，而寒冬尚未到來，出遠門不會太不便。又，為什麼選在星期二？早期有些地區人煙稀少、交通不便，選民從住家可能需要花 1 天的時間，

才能抵達投票站。而星期日是基督安息日(Christian Sabbath)，須上教堂做禮拜，隔日星期一才適合上路。所以選擇星期二為投票日。自 1845 年後，1 百多年來，雖“時空環境”與昔日農業時代早已大不相同，但投票日卻一直未再異動。

美國總統原本並無任期限制，首位總統華盛頓於當了兩任後，便不再續任，回復平民身分。他認為天下沒有不可取代的人，因此任何職位，在位者都不須以為會後繼無人，而有捨我其誰的使命感。不必戀棧，只要你願退下，多的是想取而代之者。自此美國總統最多兩任，便一直是個不成文的規定。華盛頓立下了榜樣，之後的美國總統，大都遵守任期只限兩任的傳統。但畢竟無明文規定，不見得人人皆認為該理會什麼榜樣不榜樣。曾有幾位企圖打破傳統，但一直都未成功。直到 1932 年當選的羅斯福總統，在他第二任的後期，運用了一些手腕，讓他所屬的民主黨，提名他競選第三任總統。他也選贏了，於是自 1941 年起，進入第三任。之後又贏得第四任。只是再多的手腕，都敵不過上帝的召喚。他在第四任就職(1945 年 1 月 20 日)後不久，4 月 12 日那天過世了，也結束其總統任期：從 1933 年 3 月 4 日，至 1945 年 4 月 12 日，長達 12 年 1 個多月。這是空前絕後，因從 1951 年開始生效的美國憲法增修條文第二十二條，明定“任何人被選為總統者，不得超過兩任；任何人繼任為總統或代行總統之職權者，期間如超過一任中的兩年以上，任滿後僅能獲選連任一次。”羅斯福在第 4 任任內去世，由副總統杜魯門

繼任。1948 年，杜魯門競選連任成功。因第一任當了 3 年多，超過兩年，依新規定，他不能再競選總統了。因此杜魯門的總統任期，至 1953 年 1 月 20 日結束，雖短於 8 年，但完全結束了。

歷史不長的美國，其實是個相當注重傳統的國家。就以總統選舉為例。選舉人制度的產生，乃有其時代背景。建國初期，鑑於當時除少數知識分子外，一般人對政府的運作，並不太了解，甚至全國識字率也不高，因此讓社會賢達或菁英，代表人民投票選總統，可說是一較好的方式。時至今日，民智早已大開，資訊流通又那麼便捷，這套間接選舉的機制，卻仍然徒具形式地維持。其他諸如選舉年、投票日及就職日等細節，雖皆曾變動過，但以兩百多年而言，變動算是極少。而 4 年一任，可從未修訂過。美國從華盛頓總統算起，現任拜登(Joseph Robinette Biden Jr.，1942−)是第幾屆總統？如果有人好奇，只要先求出 2020-1788=232，再將 232 除以 4，得 58，最後 58+1=59，很快便得出拜登是第 59 屆總統。反觀中華民國建國才 1 百多年，“中華民國憲法”自 1947 年 12 月 25 日施行，1948 年選出第一任總統(及副總統)，至今不過七十餘年，但並不易快速算出目前的總統是第幾屆？

底下略說明美國的選舉人制度。首先，如前所述，選舉人團乃選舉人的總稱，是每次總統選舉投票後才組成。成員不見得固定，唯一的任務，就是投票選出總統及副總統。目前選舉人團的名額是 538 位，於總統選舉年的 11 月投票結束後，在 12 月的第 3 個星期一，分別在各州的首府，及首都

華盛頓哥倫比亞特區(Washington, D.C.，以下簡稱特區)集會，投票選舉總統和副總統。但那場投票，只是形式而已，因各州通常都會要求選舉人，宣誓保證會將票投給他所屬政黨推出的候選人。絕大多數的選舉人會遵守其承諾，但亦曾有選舉人，投給他不該投的總統候選人，這種稱之為失信選舉人(faithless elector)。至 2020 年，兩百多年來，美國共出現 158 位失信選舉人。即使如此，至今不曾發生因而改變總統選舉結果者。所以一旦普選開票結束，各候選人的選舉人票決定後，誰當選便再無懸念了。

自 1961 年起，總統候選人，必須拿到超過半數(538/2=269)，即 270 張的選舉人票才能登上總統寶座。若無一候選人達標，則由眾議院(United States House of Representatives)決定：每州的眾議員代表合投 1 票，就選舉人票較多之前 3 名，選出總統。但這種情況，歷來僅發生過 1 次。那是 1824 年的選舉，當時選舉人票共 261 張，主要候選人有 4 位。亞當斯(John Quincy Adams，1767−1848)得 84 票，排名第 2，最高票有 99 票。但亞當斯逆轉勝，幸運地經眾議院選出為美國第 10 屆總統(1825−1829)。亞當斯的父親即另一位亞當斯，乃開國元勳，接任華盛頓，為美國第 3 屆(1797−1801)總統。

538 這個數字又是怎麼來的？美國國會有參議院(United States Senate)，及眾議院兩院。全美國有 50 州，每州的選舉人票，為參議員(senator)數及眾議員(Congressman)數之和。不分大小州，每州參議員均為 2 名，共 100 名。眾議員共 435

名，各州的眾議員數，乃按人口比例分配，但每州至少有 1 名。所以選舉人票，每州最少有 3 張。50 州的選舉人票，共 535(=100+435)張。另外，特區自 1961 年起，有 3 張選舉人票，因而全國總共有 538 張選舉人票。美國首都的公民，居然曾長期只能當總統選舉的旁觀者，真是咄咄怪事。原本專為預測總統大選而成立的 538 網站(FiveThirtyEight)，其中的 538，就是源自 538 張選舉人票。

附帶一提，選舉人票並非一直是 538 張。之前指出，1824 年時，全美選舉人票共 261 張，1936 年時，選舉人票只有 531 張。除特區是自 1961 年起，才有選舉人票，阿拉斯加(Alaska)州於 1959 年 1 月 3 日加入美國，成為第 49 個州；夏威夷州，則於 1959 年 8 月 21 日，成為美國第 50 個州。每多 1 州，便會增加若干張選舉人票。而除內布拉斯加(Nebraska)及緬因(Maine)兩州是採用眾議員選區方式(即在每一眾議員選區的總統選舉，獲勝者各獲 1 張選舉人票，在全州總統選舉獲勝者，獲剩下的兩張選舉人票)，其餘 48 個州和特區，均採“勝者全得”制，此造成候選人有可能全國的普選得票較高，卻輸掉選舉。例如，在幾個人口多的州均小輸，卻在好幾個小州大勝，便可能贏了總票數，卻輸了選舉人票。

美國人口普查局(United States Census Bureau)，每 10 年進行一次人口普查，最近的一次是在 2020 年執行。各州人口的消長，可能會影響該州在全國的政治地位。因人口在全國中，所佔比例增加者，眾議員數及選舉人票數，便可能增

加，反之則可能減少。2020 年的人口普查後，加利福尼亞(California)、伊利諾(Illinois)、密西根(Michigan)、紐約(New York)、俄亥俄(Ohio)、賓夕法尼亞(Pennsylvania)，及西維吉尼亞(West Virginia)等 7 州，將分別失去 1 個眾議員席位；而科羅拉多(Colorado)、佛羅里達(Florida)、蒙大拿(Montana)、北卡羅來納(North Carolina)，及奧勒岡(Oregon)等 5 州，將各增加 1 個眾議員席位，德克薩斯(Texas)州則將增加 2 個席位。上一次 2010 年的人口普查後，德克薩斯州已增加 4 席。也就 20 年下來，德克薩斯州眾議員共增加 6 席，真是不得了。

2020 年的美國總統選舉，選舉人票最多的是加利福尼亞州，有 55 票。其次是德克薩斯州，有 38 票，再來是佛羅里達及紐約州，各有 29 票。此 4 州之選舉人票，合計 151 張，佔當選所需 270 票的 55.9%以上，是欲登大位者必爭之地。至於阿拉斯加、德拉瓦(Delaware)、蒙大拿、北達科他(North Dakota)、南達科他(South Dakota)、佛蒙特(Vermont)，及懷俄明(Wyoming)等 7 個人煙稀少的州，以及特區，均只有 3 張選舉人票，可能難獲總統候選人青睞。蒙大拿州於 2024 年的總統選舉，選舉人票將有 4 張，脫離 8 小州及特區 3 票的陣容。那特區的選舉人票，有可能增加嗎？並非不可能，只是依規定，特區的選舉人票，“不得超出人口最少的州之名額”，目前特區的人口約 70.2 萬，而自 2024 年起，有 3 張選舉人票的 6 州裡，現今人口最多的德拉瓦州約 96.7 萬人，人口最少的懷俄明州約 57.8 萬。這 6 州的人口，若想持續成

長到皆能有 4 張選舉人票，恐非易事。因而在可預見的將來，特區的選舉人票，應不會增加。

大小州的政治實力差很多，不難想像，各州得政府會不斷改善居住品質及經濟條件，儘量營造更好的環境，使近悅遠來，以增加人口數，進而提高本州在全國的影響力。

19 李奇曼的美國總統選舉預測

統計是執行預測之一重要工具，但預測通常是不容易次次精準的。即使是氣象局，擁有很多儀器及專家，對颱風走向之預測，也屢不準確。其他如經濟學家、政治學家，或各類大小專家，經常在媒體上所發表的種種預測，留意的人可能不會太多，因知道大部分是不準的。

投擲一出現正面機率為 0.7 的銅板，要預測出現正面或反面？“正面”當然是合理的選擇，但每 10 次便將錯 3 次左右。銅板不會作假，也不像運動員，原本較不易出現的一面，因受到某種激勵，士氣大振，而提高出現的頻率。雖如此穩定，但任你是什麼專家都沒用，只要持續投擲，便難以每次預測出現那一面皆正確。仍以前述出現正面機率為 0.7 的銅板為例，假設各次投擲的結果為獨立，則 n 次皆出現正面的機率為 0.7^n。此值隨著 n 之增大，將漸減至 0。因此雖預測正面乃合理，但誰敢保證每次都能預測正確？有人問，連續 n 次皆出現正面，n 很大時雖是小機率事件，但難道就不會發生嗎？不是曾有報導，有人連續多次(如 15、20 次)擲出聖筊(每次出現的機率為 1/2)，那又是怎麼一回事？事實上，連續 20 次皆出現聖筊的機率為 $0.5^{20} \approx 9.54 \times 10^{-7}$，約百萬分之 1，乃相當小的機率。但只要投擲的人夠多，如有 2 百萬人參與投擲，則有人辦到，便不算稀奇。我們知道，小機率事件雖不易發生，一旦碰上大樣本，其發生便不見得仍屬

罕見了。就如樂透彩中頭獎機率雖極低，但由於買的人不少，所以仍不時會有幸運兒。

2016 年，美國總統大選落幕後不久，李奇曼(Allan Lichtman，1947−)，突然收到一封信，來自剛當選總統的川普(Donald John Trump，1946−)。信封裡裝的是“華盛頓郵報”(The Washington Post)在 9 日 23 日，即大選前一個多月的一篇報導，標題是“Trump is headed for a win, says professor who has predicted 30 years of presidential outcomes correctly”，儘管川普當時在全國民調中大幅落後，李奇曼卻預測，“川普將勝選”，川普陣營應視為雪中送炭吧！只見川普在前述報導上，用馬克筆寫著“教授-恭喜-你說對了”(Professor-Congrats-good call)，還簽上了他有如心電圖般的標誌性簽名。

李奇曼何許人也？他是美國美利堅大學(American University)的歷史系教授。依一般媒體的報導，從 1984 至 2020 年，他已“正確預測”美國 10 次總統的選舉結果。這當然很厲害，連大公司蓋洛普也做不到。2016 年 9 月，李奇曼公佈川普將贏得 11 月選舉的預測。此預測夾在當時眾多對川普嗤之以鼻的報導中，可能並沒引起太多人在意。但空谷足音，川普陣營當然注意到了，視為先知，選後遂去信捧他一下。李奇曼並非川普粉絲，選後他繼續公佈其預測。他說川普將會被共和黨主導的國會彈劾，總統一職則由副總統潘斯(Mike Pence，1959−)接任。此大膽的預測，並未成真，川普的確被提出彈劾，啟動的是民主黨，且未成功，因而潘斯並

沒機會登大位。2020 年 9 月，李奇曼再度預言，民主黨推出的拜登(Joseph Robinette Biden Jr.，1942–)，將打敗尋求連任的川普。他預測對了，但這次選後川普當然不再寫信給他了。

李奇曼如何預測？雖許多選舉的預測是基於民意調查，但對一個教授，並無法動員大量人力去做民調，那他怎能一再準確預測？在李奇曼所著的“預測下一任總統：白宮之鑰”(Predicting the Next President: The Keys to the White House)一書中，他利用 13 個關鍵指標(keys)，也就是他所說的歷史因素(historical factors)，其中有 4 個政治因素、7 個表現和 2 個性格(four political, seven performance, and two personality)，以決定總統選舉的結果。在他列出的 13 個關鍵指標中，若至少有 8 個答案是肯定的(When eight or more of the following statements are true)，換句話說，若否定答案不超過 5 個(when five or fewer are false)，則他預測執政黨(即現任總統所屬政黨)會贏得總統選舉；否則他預測執政黨會輸掉總統選舉。在此書的 2020 年版，李奇曼並將他的 13 個關鍵指標，應用至 1860 年以來的每次總統選舉結果，當然經他檢驗後，都符合其理論。我們列出李奇曼的 13 個關鍵指標如下：

1. 期中所得：期中選舉後，執政黨在眾議院席位增加。
2. 黨內競爭：執政黨的總統提名人間，沒有強烈的競爭。
3. 尋求連任：執政黨的候選人是現任總統。

4. 第三勢力︰沒有重要的第三黨派或獨立競選人。
5. 短期經濟：競選期間經濟並未陷入衰退。
6. 長期經濟：現任政府任期內，實際人均經濟成長，不低於前兩任總統任期內的平均成長。
7. 政策變化︰現任政府對國家政策產生重大影響。
8. 社會動盪︰現任政府在任期內沒有持續的社會動盪。
9. 政府醜聞：現任政府無重大醜聞。
10. 外交和軍事挫敗︰現任政府沒有在外交或軍事上有重大挫敗。
11. 外交和軍事成功︰現任政府在外交或軍事上取得重要成就。
12. 執政黨候選人具魅力︰執政黨的候選人極具魅力或為國家英雄。
13. 在野黨候選人不具魅力︰在野黨的候選人不具魅力且非國家英雄。

由李奇曼所列出的 13 個檢驗問題，可看出他認為選民選擇下一任總統，其主要的依據，乃是對過去 4 年國家的發展情況是否滿意。若滿意便繼續支持執政黨，否則便支持換黨執政。依此觀點，李奇曼等於認為，選民大部分相當理性，競選活動對他們幾乎沒有影響，選民不會被琳瑯滿目的競選

宣傳所左右。另外，李奇曼預測所依據的那 13 個關鍵指標，應是由他本人作答。當然他不無可能會依據若干客觀的數據(如第 2 指標執政黨內是否有強烈的競爭，及第 5 指標經濟是否衰退)，或參考民調，以儘量公正地評估。但其中亦有幾道題，答案究竟是肯定或否定，其實難以避免主觀。如 2121 年 8 月 25 日，中國時報有則報導，標題是“XXX 就職周年 藍綠評價兩極”。原來某市長上任滿 1 周年，某政黨議員批評他“霸混甩茫騙不及格”，但同黨議員則肯定他“施政有目共睹給他 99 分”。13 個關鍵指標中，至少第 12 及 13，有關魅力的題目，要是由執政黨作答，答案想必都是肯定，但若由在野黨作答，答案不必說都會是否定。尤其關鍵指標之第 12 及 13，究竟什麼樣的人算有魅力？柯林頓有魅力嗎？川普有魅力嗎？拜登有魅力嗎？有人可能認為 3 人中，沒一個有魅力吧！其於諸如社會動盪、醜聞、挫敗、成功之認定，也都有些主觀。以醜聞為例，同一件事，支持者可能覺得連醜聞的邊都沾不上；反對者卻可能會視為罪大惡極。

在 2020 年 9 月的一場訪問裡，李奇曼說，至 2019 年底，只有 4 個指標對尋求連任的川普不利，但進入 2020 年後，因新冠疫情擴大，加上反種族歧視的示威，在數個月內席捲全美，對川普不利的關鍵指標，一下子增加了第 5、6 及 8 等 3 項。否定答案達到 7 個，那便意味著川普的連任不會成功，而後來川普也的確落選。雖李奇曼能依據現有數據，包含各家民調，但就憑回答那簡潔的 13 個指指標，殺牛用雞刀，對總統選舉之預測，準確率便能超越蓋洛普等資源不少

之大機構？他真的是神算子嗎？

李奇曼的揚名立萬、聲名遠播，大致是始於 2016 年的美國總統選舉，即共和黨的川普當選那次。特立獨行，望之不似人君的川普，選前不斷拋出各種歧視言論，屢引起爭議，常連與他同黨者都看不下去。愈臨近投票日，愈多擔心被拖累的共和黨要角，迅速跟川普割袍斷義。選舉想獲勝，得吸收各族群的票，宜多方討好、面面俱到。從頭到尾走偏鋒，不在乎激怒人，導致眾叛親離的川普，豈有當選可能？共和黨裡熟悉選情的專家，早早覺得選舉大勢已定，只能期待下次了。即使不落井下石，也多半作壁上觀。在此氣氛下，獨排眾議，預測川普是贏家，選後當然令人刮目相看，連川普都給他寫信了。

事實上，2016 年 11 月 8 日投票後，11 月 10 日，在網站 Boston.com 就有標題為“Here are 3 people who correctly predicted Donald Trump would win the election-And why they got it right.”之一篇報導。其中指出，除李奇曼外，另有兩位正確預測川普會當選者。一位是石溪大學(Stony Brook University)的政治學教授諾普思(Helmut Norpoth，1943−)，他使用統計的模式，分析歷屆大選，歸納出初選模型(the Primary model，指 1912 年至今，初選過程中表現較佳的候選人，通常在大選時能夠勝出)，及鐘擺效應模型(Swing of the Pendulum model，指自 1960 年至今，政黨連續兩任入主白宮的機率為 54.7%，之後再繼續執政的機率下滑到 49.4%)。他成功預測過去 5 次勝選總統，而在投票前 9 個月，他就預測

川普有 97%的機率會當選。另一位是美國知名的紀錄片導演摩爾(Michael Francis Moore，1954–)，他給出 5 個川普會贏的理由(5 Reasons Trump Will Win)。所以，李奇曼倒也非唯一的先知。另外，前面說過，從 1984 至 2020 年，李奇曼正確預測美國 10 次總統的選舉結果。但所謂正確預測，究竟是預測什麼正確？美國總統選舉時，乃採複雜的“勝者全得”(只有兩個州不是)制，不用統計、不必民調，怎有辦法次次成功預測誰當選？

2000 年的總統選舉，在 13 個關鍵指標中，執政的民主黨，肯定答案有 8 個。民主黨的候選人，時任副總統的高爾(Albert Arnold Gore, Jr.，1948–)普選得票率較高，較對手多了約 0.51%，贏了 54 萬餘票，但他的選舉人票僅 266 張，不足當選所需之 270 張，共和黨的布希(George Walker Bush，1946–，常被稱為小布希(Bush Junior)，因他父親老布希(Bush Senior，George Herbert Walker Bush，1924–2018)亦擔任過美國總統)當選。李奇曼辯白說，他預測的是普選獲勝者(popular vote winner)。在李奇曼 1988 年出版的“總統的 13 把鑰匙(The Thirteen Keys to the Presidency)一書中，他的確將其模型定義為預測普選結果。只是李奇曼在期刊發表他對 2000 年的總統選舉之預測時，並未在其文章中，提醒讀者這種細微差別(nuance)，他就僅是預測高爾會贏(He simply predicted that Gore would win)。我們說過了，贏得普選是有可能輸掉大選。但自 1888 年以來，到 2000 前，那 1 百多年間，美國可沒有發生過這種普選及大選贏家，分屬不同人的情既。因而看到

李奇曼之預測文章者，應很少人不是解讀成李奇曼預測高爾當選。話說回來，對總統選舉之預測，豈有不是預測誰當選者？

2016 年的總統選舉，13 個關鍵指標中，執政的民主黨，肯定答案僅有 6 個，因而李奇曼預測挑戰者川普將是普選獲勝者。開票後，民主黨的候選人柯林頓(Hillary Diane Rodham Clinton，1947–)普選得票率較高，比對手川普多了約 2.1%，多達 286 萬票餘。李奇曼的預測再度錯了嗎？沒有！李奇曼宣稱，鑑於選舉團與普選之間的差異急劇增加，自 2000 年的總統選舉後，他不再預測普選結果，而只是預測誰當選總統。有趣的是，13 個關鍵指標並未修改，卻能左右逢源？也幸好李奇曼改變其預測目標，否則 2016 年他的預測便錯了，因普選獲勝者是柯林頓。但柯林頓雖贏得普選，選舉人票卻僅有 227 張，由得到 304 張選舉人票的川普當選總統。川普也成為美國歷來第 5 位總統當選人，但輸掉普選者。只是 227+304=531？不足 538，怎麼回事？原來有 2 名該投給川普的共和黨選舉人，及 5 名該投給柯林頓的民主黨選舉人，他們既不投柯林頓也不投川普。共有 7 名失信選舉人，是 1872 年以來首次出現超過 1 名失信選舉人。2000 至 2020 年的 6 次選舉，除 2016 年外，2000 年及 2004 年，各有 1 位失信選舉人。

其實，就算沒有什麼關鍵指標，從下次 2024 年的美國總統選舉起，至 2060 年，均以投擲一銅板來預測，正面就預測執政黨的候選人當選，反面則預測執政黨的候選人落

選。能 10 次正確預測美國總統的選舉結果，並不容易，機率也不見得是 1/1,024(即使銅板為公正，為什麼？)。但只要有幾千人這樣做，則 2060 年，若有人宣稱他有一“神奇銅板”，連續 10 次美國總統選舉正確預測，便不足為奇了。

20 美國總統選舉預測不易

由於 1888 至 1996 年，那 1 百多年間，美國總統選舉結果，全是由普選得票率較高者當選，人們習以為常，所以一般做民調都是預測普選得票率，而誰會當選？想當然耳就是被預測得票率最高的那位。在那段時期，普選得票率最高，等同於當選。蓋洛普公司於 1935 年成立，自 1936 至 2012 年間，歷經 20 次總統選舉，他們的預測誰當選，有 3 次錯誤，分別是 1948、1976，及 2012 年(這 3 次皆由普選得票率較高者當選)；2000 年預測誰當選正確，但他的普選得票率卻較低，因而嚴格講這次並不算預測正確。2012 年預測錯誤後，蓋洛普便宣佈自此，不再對美國總統選舉進行預測，而致力於其他議題的民調。畢竟在民調時，對總統選舉回答會投 A 者，實際卻可能投 B，甚至可能根本就沒去投票，實在太難預測了。在“世説新語”“雅量篇”裡，周顗(269−322)指著州官顧和(288−351)的心問他，“此中何所有？”顧和答以，“此中最是難測地。”約 1700 年前晉朝時的顧和，便了解人心難測！僅憑回答幾道題目，實在無法精準掌握被問者之投票動向，造成預測困難。但主要是，美國各州(特區)的選舉人票，多寡差異極大，而“勝者全得”的選舉制，造成有可能贏得普選卻輸掉大選。當雙方支持度很接近時，只要一州翻盤，便可能使選舉翻盤。

在 2016 年的那次總統選舉，共和黨提名的川普，口無

遮攔、信口開河，更被認為有逃稅記錄。但他砲轟政府、攻擊對手之激進言詞，卻讓許多不滿現況的美國人聽了很爽。他那句“Make America Great Again”之競選口號，有如昔日台灣某候選人提出的“有夢最美，希望相隨”，讓不少美國人，相信川普能帶領美國，恢復昔日榮景。在競選期間，川普一再打破政治慣例，很多主流媒體及傳統政治人物都不喜歡他。反觀民主黨推出的候選人柯林頓，曾任 12 年州長夫人、8 年第一夫人、8 年參議員，及 4 年國務卿，30 多年的公職生涯，從政經驗豐富，論述能力、專業形象與繳稅紀錄，縱有零星爭議，但條件仍明顯優於川普。選前民調及媒體，普遍看衰川普，早早不認為川普有任何當選機會的，包括共和黨的一些領袖人物。投票日該晚持續開票時，當川普的普選票數落後愈來愈多，選舉人票卻領先愈來愈多，眼看川普就要當選，共和黨的一重要策士墨菲(Mike Murphy，1962−)，在其推特(Twitter)上寫著：

> I've believed in data for 30 years in politics and data died tonight. I could not have been more wrong about this election.

數據何用？從今晚起，數據死了。一輩子崇尚數據、歷來依數據做決策者，在說出數據已死時，心裡想必悲痛不已。而以預測總統選舉起家的蓋洛普，當宣佈放棄日後的總統選舉預測時，其心境應有如傳說中的上古射日英雄后羿(生卒年不明)，當他停止用弓時，應也很悲痛吧！只是數據真的無用嗎？或者說，數據到底何用？

“時代”(Time)雜誌於每年底，都會選出年度風雲人物(Person of the Year)。毫無意外，2016 年選出的是川普。自 2013 年起，便擔任“時代”雜誌主編(Managing editor)的吉布斯(Nancy Reid Gibbs，1960−)說：“我們何時見過一個人，公然違抗所有期望、打破規範、違反準則，在百分之 1 的獲勝機率下，卻贏得選舉。在那過程中，他不只擊敗一個，而是兩個政黨。”即使川普當選了，即使他被選為年度風雲人物，“時代”雜誌對他仍沒什麼好評。只是選前川普的獲勝機率，真的只有百分之 1 嗎？如果連高度掌握資訊的“時代”雜誌，直到選後 1 個月，都仍以為川普當選的機率那麼低，一般人(包括前述共和黨的策士墨菲)在選前，自然更不會認為川普有任何獲勝的可能。見到民主黨的柯林頓聲勢浩大，而川普卻經常被媒體圍剿，情勢一面倒，會讓人以為大局已定，柯林頓即將入主白宮。數據固然會說話，但若選前未了解數據在說些什麼，則選後覺得統計騙人，其實是被自己騙。這點我們之後將會說明。

2016 年 11 月 29 日，中國時時報有一則報導，標題為“密西根完成美總統大選計票川普險勝”，指出密西根州於 11 月 28 日，投票通過認證總統選舉計票結果，川普領先柯林頓 1 萬多張票，密西根州的 16 張選舉人票，全由川普拿下。密西根州是當年最後一個確認總統選舉結果的州。11 月 8 日就投完票，要花 20 天，才能確認結果？這其實一點都不稀奇。2020 年的美國總統選舉，11 月 3 日就投完票，但密西根州於 11 月 23 日，再度是 20 天後，才通過認證選舉結果。這

次換成民主黨的拜登在該州領先川普 15.4 萬票，取得該州的 16 張選舉人票。不過這次密西根州可沒殿底，最後一個確認總統選舉結果的州，為西維吉尼亞州，12 月 9 日才確認，由川普拿下該州 5 張選舉人票。而總統交接時，負責向候任總統團隊，移交聯邦政府廳舍使用權的美國聯邦總務署(General Services Administration)署長，亦於 11 月 23 日，正式認證拜登為新任總統，川普政府自此才展開交接程序。總務署的認證看起來很遲緩，有如在偏袒不認輸的川普，但想想那時仍有幾個州，尚未確認總統選舉結果呢？

何以美國的總統選舉，各州對選舉的確定，要這麼久？

台灣公職人員選舉目前的投票流程，著眼於防弊，至今可說沒什麼彈性。美國總統選舉投票日在 11 月，若因在外地工作或就學等原因，可採不在籍投票(absentee voting，亦稱缺席投票)，可能從 9 月或 10 月便開始了，屬於提前投票(Early voting)的一種。除不在籍投票(absentee voting)外，提前投票尚包含郵寄投票(mail-in voting)，及提前親自到投票所投票，各州規定不同。2020 年，因新冠肺炎疫情的關係，為避開投票日當天投票所的排隊人潮，郵寄選票大量增加，好幾個州還修改法規，以便民眾寄出的選票在 11 月 3 日總統選舉投票日後之幾天內還可收件。這也使得計票過程延後。

2016 年的美國總統大選，總投票數約 1.357 億，估計有超過 5 千萬，即差不多 4 成的投票人，選擇提前投票。2020 年由於疫情關係，提前投票的選民人數，更創下歷史新高，

總計約有 1.032 億人，佔總投票數約 1.583 億的 6.5 成。提前投票人數多，表示計票時間將延長，有些州允許提前清點提前投票，有些州則必須在投票日當天，才能計算提早投的票。另外，美國有各式各樣的投票工具，不限紙張選票，觸控式投票、語音投票、槓桿投票機，及機器掃描選票等。每個地方的投票裝置也不太一樣，有些州還提供不同語言的選票。而各州對不在籍投票的規定，也不盡相同，有些州甚至允許已提前投票的選民，能反悔更改選擇。各州更改投票及重新投票的規範也有不小差異。2020 年，爭取連任的總統川普，在最後關頭還呼籲，已投票的選民改投他，為複雜的選情增添變數。共和黨籍的川普，向來懷疑郵寄投票，存在選舉舞弊之風險，因此共和黨選民較不傾向郵寄投票。在 11 月 3 日投票日當天，預估投票者中，以共和黨選民居多。但有些州先計算郵寄選票，有些州先計算親自投票的票數，因而開票時，可能會呈現不平均的走勢，兩大黨陣營的任何一方，在開票夜得票領先的現象，可能都只是暫時性的，所謂“紅色幻影”(Red Mirage)，或“藍色幻影”(Blue Mirage)，其中紅色及藍色，分別是共和黨及民主黨的代表色。因而唯有全部選票都開完，才知誰是最後笑的人。開票夜東部時間 11 月 4 日凌晨約 2 時 30 分，川普在白宮向支持者發表演講，宣稱自己贏得了大選，呼籲停止一切計票活動，並說繼續計票是對美國人民的欺詐。對此，拜登陣營譴責川普的行為，拜登則敦促其支持者在計票時保持耐心，並相信自己將贏得這次選舉。在等待結果時，人心的焦慮及煩躁可以想像。最後拜登及川普的普選得票率，分別是 51.3%，及 46.9%，拜

登贏了 705 萬多票，選舉人票則兩人分別有 306 及 232 張。

不在籍投票，可透過郵寄、電子郵件、傳真，或到指定地點投票。沒有簽名的郵寄選票，不會被計入，但若發現得早，尚可補救。這些林林總總的投票方式，的確讓選民方便許多，也相當有人性。往好處想，能提高投票意願，但卻也因而增加計票的複雜度。對有效票的認定，也時有爭議。當把選票投進具有光學掃瞄功能的投票機中，機器會自動判讀票上的畫記內容。若投票機讀到無效標記，會退出選票，此時可要求換一張新的選票卡。這過程中會有弊病嗎？而投票的管道那麼多，所有票都有拿出來計數嗎？確定一張都沒漏掉？仰賴電子郵件及電子投票器，會有駭客入侵嗎？票數會被竄改嗎？要知當兩主要候選人的票數差異較大時，可能還無所謂。像 2016 年，雖至 11 月 28 日，各州才全部通過認證總統選舉的計票結果，但在 11 月 9 日，即總統選舉投票日(11 月 8 日)之隔天，柯林頓便承認敗選了。這是因採“勝者全得”制，若在某州有些微的誤差，並不至於影響該州選舉人票之歸屬。柯林頓在政壇久矣，歷經大風大浪，數數選舉人票，就大方認輸，留下漂亮的下台身影。但當雙方票數很接近時，便得反覆確認是否計算正確。這可能是有些州要花那麼多天，才能確認選舉結果的主要原因。以美國幅員那麼大，各州在選後多日，仍陸續“發現”一些選票，長期以來，恐怕常是如此，美國人應早就見怪不怪了。只是若遇到不乾脆的候選人，是可一再糾纏。如 2020 年的川普，不但在好幾州要求重新計票，甚至提起幾十件質疑選舉結果的訴訟，

但終究無法翻案成功。以亞利桑那州為例。此州一向支持共和黨，從 1952 年到 2016 年的 17 次美國總統選舉，共和黨候選人在此州贏了 16 次。但 2020 年拜登卻在該州以 0.30%、1 萬餘票的差距險勝川普，也讓川普與川粉編織各種藉口要求重新計票。由於亞利桑那州是共和黨主政，有執政優勢，因此相關重新計票工作，從 2021 年 3 月起展開。花了 6 個月，2021 年 9 月 24 日報告出爐，不但找不到弊端，且讓拜登的得票多了 360 張。早有理論顯示，在無作票的先決條件下，重新計票將使原領先者，得票更多。

由於不在籍投票制，當候選人仍馬不停蹄地四處競選時，卻有不小比例的選民，早已投完票作壁上觀了。據估計，提前投票者，支持民主黨的比例較支持共和黨的高。提前投票究竟對那一黨有利？散在世界各地的美國公民(不包括軍人)，總數超過 800 萬，美國約有 40 個州的人口還沒這麼多呢！他們的政黨傾向及投票行為，與國內選民，相差很大嗎？選情變化萬千，國外選民，都能有效掌握國內訊息嗎？美國國土橫跨 6 個時區，各地投票時間也不一致，當有些地區已開完票，還在投票的地區，會受到影響嗎？這些因素，再加上“勝者全得”制，使美國總統選舉的預測，不能想像會有神算子存在的可能。

2020 年的美國總統選舉，拜登得 306 張選舉人票，川普得 232 張，拜登當選。其中得票率之差異小於 1%者(底下括號內皆顯示得票率之差異)，有喬治亞(Georgia，0.24%)、亞利桑那(Arizona，0.30%)，及威斯康辛(Wisconsin，0.63%)等

3 州。這 3 州都是拜登領先，因而總共的 37 張選舉人票，全歸拜登。一旦少了這 37 張票，拜登的選舉人票，將降至 269 張，便未超過半數 270 張，則今天川普便能繼續大放厥詞了。以李奇曼那 13 道題，豈有辦法區隔這麼微小的差別？另外，賓夕法尼亞州(1.16%)有 20 張選舉人票，歸拜登；北卡羅來那州(1.35%)，有 15 張選舉人票，歸川普。此二州之差異，當然也都很小。直到開完票前，前述這兩州的選舉人票將落入那一陣營，誰能斷言？蓋洛普公司大唱歸去來兮，退出美國總統選舉之預測，是可理解的。

21 數據已死？

2016 年的美國總統選舉，身為共和黨策士的墨菲，坐擁數據，羽扇綸巾，因而從來不看好川普。11 月 8 日開票當晚，眼看川普就要當選，不禁慨嘆“數據已死”。那 2020 的美國總統選舉後，有沒有人也慨嘆些什麼呢？2020 年投票日(11 月 3 日)的兩天後，11 月 5 日，“紐約時報”(The New York Times)有一則報導，標題為，“2016 年對投票造成打擊，那 2020 年有毀了它嗎？”(2016 Dealt a Blow to Polling. Did 2020 Kill It?)此文一開始便指出，“選舉之夜，計票開始後不到 24 小時，媒體和政治圈已經宣布一個失敗者，那就是民調(polling)。”即雖尚不知那一候選人會落選，但已確定的是民調輸了。然後引述一些民調專家的意見，包括“對我們這一行來說就是完了”；“民調似乎已崩潰了，救不回了”；“民調業就是一堆廢棄物，炸掉算了。”看來經過 4 年，民調落入比“數據已死”更淒慘的狀態。

2016 年的美國總統選舉，選前各家民調，對柯林頓普選得票率領先的預測，平均略低於 4%，最終她贏了約 2.09%，誤差並不算太大，與 1968 年以來民調的平均誤差差不多。因而許多民調專家，並不認為他們的預測太離譜。只不過柯林頓雖贏得普選，卻輸掉大選。因美國的總統選舉，並非由普選得票率之多寡來決定，而是由各州的選舉人票，尤其是搖擺州。在總統選舉時，搖擺州向來為兩大黨極力爭取的目

標。像美國有 18 個州及特區，自 1992 至 2012 的總統選舉，都是投給民主黨的候選人，是所謂民主黨的藍牆(Blue Wall)。2016 年風雲變色，密西根、賓夕法尼亞，及威斯康辛等 3 州，倒向川普，這 3 州共有 46 張選舉人票，若非輸掉這 3 州，柯林頓便當選了(她應獲 232 張選舉人票，但有 5 位失信者，實際得 227 張，227+46=273>270)。2020 年拜登光復這 3 州，分別贏 2.78%、1.17%，及 0.63%，若沒這 3 州，拜登便落選了(他獲 306 張選舉人票，306-46=260<270)。選前眾人皆知，拜登只要保住柯林頓 4 年前勝出的州，加上讓此 3 搖擺州搖回來，得票便超過入主白宮所需的 270 票。拜登達成目標，且翻轉的還不只上述 3 州。共和黨的鐵票倉亞利桑那州，自從 1952 年開始，除 1996 年外，此州的總統選舉，全由共和黨勝出，但拜登以驚險的約 0.30%得票率領先攻下，11 張選舉人票入袋；另外，過去 6 屆的總統選舉，喬治亞州均由共和黨候選人以大差距拿下，此次拜登贏了約 0.24%，是不太多，但已夠將紅土染藍了，又是 16 張選舉人票。

2020 年的總統選舉，最後雖如大部分的預測，拜登當選，但在州這級，則有好幾州的誤差不小。如民調預測，在佛羅里達州，拜登約領先 2.5%，結果倒輸了約 3.3%；在俄亥俄州，川普領先不到 1%，結果大贏約 8.03%。更不要說在眾議院，民主黨原本就領先共和黨 35 席，許多人預測藉由擊敗川普的旋風，民主黨將再增加 5 至 15 席，結果卻是跌破眼鏡的民主黨減少 10 席，共和黨增加 16 席。民主黨雖仍

保住眾議院之控制權，但差距大幅縮小。這是從事民調者，對自己這行徹底失望的一主要原因。

FiveThirtyEight，亦簡單寫成 538，此網站於 2008 年 3 月 7 日，由西爾弗(Nate Silver，1978−)所創。網站名稱源自於美國總統選舉時，共有 538 位選舉人。自 2010 年 8 月起，此網站隸屬“紐約時報”。又自 2013 年 7 月起，網站被 ESPN 收購。為何知名媒體，會爭相將 538 網站納入麾下？因此網站，以精準預測 2008 及 2012 年的美國大選結果而出名。媒體當然樂意與如此聲譽卓著的網站結合。今日該網站所公佈的預測，涵蓋民意調查分析、政治、經濟、科學、流行文化，與體育等，早已不侷限在選舉。

為讓大家對當選機率的涵義有所了解，先舉一簡單的例子。假設美國只有 A、B、C 3 州，各有 10、15、20 張選舉人票，共 45 張，且皆採“勝者全得”制。則任一總統候選人想勝選，須拿到至少 23 張選舉人票。現設某次總統選舉，有 M 及 N 兩位主要的候選人。又設預測 M 在 3 州的獲勝機率分別為 0.5、0.7、0.8；N 在 3 州的獲勝機率分別為 0.5、0.3、0.2。看到這樣的數據，不少人會解讀成，就是 M 了，N 豈有機會？底下來計算兩人的勝選機率。由於得 23 張以上的選舉人票，表至少要贏 2 州，因此兩位候選人，要當選都有相同的 4 種可能：

(A 州勝，B 州勝，C 州勝)，
(A 州勝，B 州勝，C 州敗)，

(A 州勝，B 州敗，C 州勝)，
(A 州敗，B 州勝，C 州勝)。

又假設各州的投票結果相互獨立。由此得

P(M 當選總統)

=0.5×0.7×0.8+0.5×0.7×0.2+0.5×0.3×0.8+0.5×0.7×0.8

=0.28+0.07+0.12+0.28=0.75。

立即可得 P(N 當選總統)=1-0.75=0.25。N 雖看起來贏面不大，但仍有 0.25 的機率。這種事件，離被視為不可能發生還遠得很。因此若發生何須訝異？

若氣象局預測降雨機率為 0.25，你自己判斷不會下雨，遂不帶傘出門，結果淋了一身雨。雖懊惱，能就此指責氣象局的預測不準嗎？除非你已長期收集數據，預測不準的證據確鑿，否則顯然不行。因 1/4 的機率，表示約每 4 次，便會發生 1 次。機率用在對氣象之預測，一般人似較易理解其涵義。不會輕易慨嘆“氣象局該關門了”。

538 網站對 2016 年美國總統選舉的兩位主要候選人，表 1 給出預測及實際得票率，可看出誤差並不算大；表 2 給出當選機率之預測，川普雖不被看好，預測仍有 0.286 的當選機率，大於 1/4 的機率，比“時代”(Time)雜誌宣稱的川普僅“百分之 1 的獲勝機率”大多了。川普幸運地當選了，豈能因此導致“數據已死”？表 3 給出預測及實際選舉人票，

這部分差異較大，並且誰當選也預測錯了。但我們已多次強調，美國總統選舉，特殊的“勝者全得”制，使得當兩候選人實力相差不大時，究竟誰的選舉人票會較多，是很難精準預測的。

我們再度來討論獲勝機率，先看一簡單的例子。假設 D 君投擲一正面出現機率為 0.88 的銅板 56 次(此處並不必假設各次的結果為獨立)。因 56×0.88=49.28，即預期將出現 49 個左右的正面。不會因正面出現機率的機率 0.88 相當大，就預期可得 56 個正面。而且正面出現的次數，若在 49 附近，也都會覺得合理。事實上，若出現過多正面，可能會啟人疑竇，懷疑正面出現的機率應大於 0.88。其次考慮較一般的情況。假設有 56 個銅板，正面出現的機率不盡相同。今各投擲 1 次(仍不必假設各次的結果為獨立)，由於期望值有線性性質，因此將各銅板正面出現的機率相加，便是共得的正面數之期望值，也是合理的預測值。美國 50 州、特區、緬因州 2 個選區，加上內布拉斯加 3 個選區，總統選舉時，可視為共有 56 個採“勝者全得”的州(區)。欲估計總統候選人所得的選舉人票，便要對 56 個州(區)，皆估計獲勝機率。表 4 便是 538 網站對 2016 年美國總統選舉的兩位主要候選人，勝選機率之預測。雖候選人有多組，但因“勝者全得”，使得只有柯林頓與川普，才有獲選舉人票的可能。我們想回答一個令人感興趣的問題：州(區)預測獲勝機率較大的候選人，確實也在該州(區)贏，這種州(區)數的期望值為何？

上述問題此本質上正是之前的投擲銅板問題。對 56 個

州(區)，將預測獲勝機率較大的那位候選人(柯林頓或川普)，視為銅板正面。要知所謂正面或反面，不過有如名字，不同的銅板，指定那一面為正並無妨。如此一來，問每州(區)預測獲勝機率較大者是否獲勝，便有如問投擲一銅板正面是否出現。而問共有幾個州(區)，由預測獲勝機率較大的候選人當選，便相當於求投擲 56 個銅板，共得的正面數。將表 4 中 56 個州(區)，各取較大的預測獲勝機率，全部相加，得總和 49.35。其中為了簡便，機率大於 0.999 者，以 1 計，而機率小於 0.001 者，則以 0 計。即得由 538 網站的預測，預期有 49.35 個州(區)，由獲勝機率較大的候選人，贏得該州(區)。實際吻合幾個？50 個！這還能說不準嗎？還能慨嘆“數據已死”嗎？

若某州(區)預測獲勝機率較大者落敗，我們不妨從俗稱之為“逆轉”州(區)。由表 4 知，56 個州(區)中，共有 6 個“逆轉”州(區)。表 5 給出 538 網站，在那 6 個逆轉州(區)，對柯林頓及川普，所預測之得票率，實際得票率亦附上以為比較。6 個逆轉州(區)，得票率預測領先的都是柯林頓，分別領先 0.6%、0.3%、4.2%、0.7%、3.7%，及 5.3%。其中 0.6%、0.3%，及 0.7%，都很小，即這 6 州(區)裡，至少有 3 州(區)，兩候選人的支持率差異，應是在誤差範圍內，因而柯林頓被逆轉，並不足為奇。回過頭來看川普，他在 6 個逆轉州(區)的實際得票率，分別高出柯林頓 1.20%、10.47%、0.23%、3.66%、0.73%，及 0.75%。這些差距，除了緬因州第 2 選區(只有 1 張選舉人票，影響很小)的 10.47%，及北卡羅來納州

的 3.66%較大外，其餘在佛羅里達、密西根、賓夕法尼亞，及威斯康辛等 4 州，差距分別是 1.20%、0.23%、0.73%，及 0.75%，可說都相當小。只要些微的波動，輸贏便可能換人。這 4 州，共有 75 張選舉人票。若少掉這 75 票，川普便只有 231 票，將黯然神傷地落選。能這麼驚險的逆轉這 4 州，說川普運氣好，應不為過。話說回來，有多少預測能掌握這麼小的誤差？

其實，還不必上述 4 州全輸掉，因川普得到的選舉人票，比當選所需的 270 票，多出 36 票，故佛羅里達州與密西根州(共 45 票)、佛羅里達州與賓夕法尼亞州(共 49 票)、佛羅里達州與威斯康辛州(共 39 票)，或密西根州、賓夕法尼亞州，與威斯康辛州(共 46 票)，只要在此 4 組州的任一組輸掉，則“時代”雜誌 2016 年選出的風雲人物，便不是川普了。

事實上，我們已數度強調，若真見到次次正確預測銅板那一面朝上的人，人們恐怕不但不信他，反而懷疑其中有假。所以如何能對這麼複雜的總統選舉，要求處處精準預測？

這是一雙方實力接近的總統大選，“勝者全得”制，造成正確預測誰當選之困難。若如一般選舉，依得票率的高低，以決定誰當選，則 538 網站便有近乎完美的預測。由表 2，預測柯林頓的得票率為 48.5%，領先川普的 44.9%，實際果然柯林頓的得票率較高。而且柯林頓的實際得票率為 48.07%，比預測值僅略少 0.43%；川普的實際得票率為

45.99%，比預測值不過高出 1.09%。這樣的誤差，在選舉預測裡，算是高度準確了。只要想，從一袋中，以隨機取球來估計紅球所佔比例。兩次試驗所得比例之差異，若只有 1.09%，豈會不滿意？差異若小至 0.43%，難道不會驚嘆不已嗎？

只要了解隨機性，便知預測之誤差難免。美國總統選舉誰輸誰贏之預測不易，統計就是做到它能做的。數據不死，也未逐漸凋謝。

表 1. 538 網站 2016 年美國總統大選預測及實際得票率

候選人	柯林頓	川普
預測值	48.50%	44.90%
實際值	48.07%	45.99%

表 2. 538 網站 2016 年美國總統大選當選機率之預測

候選人	柯林頓	川普
預測值	0.714	0.286

表 3. 538 網站 2016 年美國總統大選預測及實際選舉人票

候選人	柯林頓	川普
預測值	302.2	235.0
實際值	232	306

表 4. 538 網站 2016 年美國總統大選各州(區)獲勝機率之預測

		候選人		
州(區)	選舉人票	柯林頓	川普	實際獲勝者
Alabama	9	<0.001	>0.999	川普
Alaska	3	0.235	0.764	川普
Arizona	11	0.334	0.666	川普
Arkansas	6	0.004	0.996	川普
California	55	>0.999	<0.001	柯林頓
Colorado	9	0.775	0.224	柯林頓
Connecticut	7	0.973	0.027	柯林頓
Delaware	3	0.915	0.085	柯林頓
District of Columbia	3	>0.999	<0.001	柯林頓
Florida*	29	0.551	0.449	川普

Georgia	16	0.209	0.791	川普
Hawaii	4	0.989	0.011	柯林頓
Idaho	4	0.009	0.990	川普
Illinois	20	0.983	0.017	柯林頓
Indiana	11	0.025	0.975	川普
Iowa	6	0.302	0.698	川普
Kansas	6	0.027	0.973	川普
Kentucky	8	0.004	0.996	川普
Louisiana	8	0.005	0.995	川普
Maine	2	0.826	0.173	柯林頓
District 1	1	0.915	0.085	柯林頓
District 2*	1	0.509	0.490	川普
Maryland	10	>0.999	<0.001	柯林頓
Massachusetts	11	>0.999	<0.001	柯林頓
Michigan*	16	0.789	0.211	川普
Minnesota	10	0.850	0.150	柯林頓
Mississippi	6	0.022	0.978	川普
Missouri	10	0.039	0.961	川普
Montana	3	0.041	0.959	川普
Nebraska	2	0.023	0.977	川普

District 1	1	0.107	0.893	川普
District 2	1	0.442	0.558	川普
District 3	1	0.008	0.992	川普
Nevada	6	0.583	0.417	柯林頓
New Hampshire	4	0.698	0.302	柯林頓
New Jersey	14	0.969	0.031	柯林頓
New Mexico	5	0.826	0.172	柯林頓
New York	29	0.998	0.002	柯林頓
North Carolina*	15	0.555	0.445	川普
North Dakota	3	0.023	0.977	川普
Ohio	18	0.354	0.646	川普
Oklahoma	7	<0.001	>0.999	川普
Oregon	7	0.937	0.063	柯林頓
Pennsylvania*	20	0.770	0.230	川普
Rhode Island	4	0.932	0.068	柯林頓
South Carolina	9	0.103	0.897	川普
South Dakota	3	0.061	0.939	川普
Tennessee	11	0.027	0.973	川普
Texas	38	0.060	0.940	川普
Utah	6	0.033	0.832	川普

Vermont	3	0.981	0.019	柯林頓
Virginia	13	0.855	0.145	柯林頓
Washington	12	0.984	0.016	柯林頓
West Virginia	5	0.003	0.997	川普
Wisconsin*	10	0.835	0.165	川普
Wyoming	3	0.011	0.989	川普
註：*表由獲勝機率預測值較低者獲勝				

表 5. 538 網站 2016 年美國總統大選 6 逆轉州(區)預測及實際得票率

		538 網站預測之得票率		實際得票率	
逆轉州(區)	選舉人票	柯林頓	川普	柯林頓	川普
Florida	29	48.1%	47.5%	47.82%	49.02%
Maine (District 2)	1	45.6%	45.3%	41.06%	51.53%
Michigan	16	48.4%	44.2%	47.27%	47.50%
North Carolina	15	48.2%	47.5%	46.17%	49.83%
Pennsylvania	20	48.9%	45.2%	48.02%	48.75%
Wisconsin	10	49.6%	44.3%	46.44%	47.19%

22 統計相關

某作家新近出了本書，這本書值得看嗎？現今一般人常不假思索，便進入搜尋引擎 Google 查查。立刻出現很多相關的資料。看起來評價相當不錯，決定要買後，進入某網路書店找到該書，結果出現一句“瀏覽此書的人，也瀏覽…”。所列出的書，有些你毫無興趣，但也有幾本引起你的注意。網路時代，人們若對某事物或事件有興趣，常會主動或被動，接觸相關的事物或事件。

相關就是有關係，有各式各樣的關係。數學裡的命題“若 A 則 B”，便是一種因果關係。只要某條件(或者說前題)A 成立，便導致結果 B 必然成立，這是一很強的關係。數學裡的定理，通常便是顯示某重要的因果關係成立。因果關係外，有些關係就不見得那麼強烈。事實上，只要夠努力去追究，總會找出兩物或量之間，有某種程度的關係，當然有可能很微弱。正如有些人對於食品、農藥及汙染等，無法容忍其中有某特定成分存在，要求該成分須“零檢出”。但也有專家指出，凡儀器不夠好，無法驗出的，便是“零檢出”。此隱含只要儀器夠精密，很多成分即使微乎其微，也能驗出。統計裡兩量間有所謂“相關”(correlation)，又稱為相關性，並以相關係數(correlation coefficient，有時只稱 correlation)來計量兩變數共同變化的程度。兩量相關並不表有因果關係。另外，有相關便也有“無相關”(uncorrelated，或稱 no correlation)，但無相關也不表兩量間，一定沒有關係，關係

要找總是有的。

2021 年 6 月 29 日聯合報有則報導，標題是“NBA／喬治好表現重新證明自己 卡森斯：給他該有的稱讚”。這是對西區快艇隊，在一場季後賽獲勝後的評論。快艇隊的球員莫里斯(Marcus Morris Sr.，1989−)，在該場比賽中，16 投 9 中攻下 22 分，命中率有 56.25%，算是不錯。記者誇獎他，“本季季後賽,當莫里斯命中率高於 5 成時,快艇 8 勝 0 敗。”數據如此，那可否說“若莫里斯命中率高於 5 成時，則快艇隊必勝。”假設此因果關係成立，快艇隊當然得善用，因已進入季後賽，每一場勝利都很關鍵。可惜此因果關係很可能不成立。在職業球賽裡，這類應被歸於“有趣事件”的還不少。如“xxx 隊本球季在主場出賽，若 xxx 與 xxx 合得至少 60 分，則 xxx 隊至今尚未敗過”。30 支球隊，又那麼多場比賽，仔細觀察，總會發現一些有趣的現象。其中大部分，就只是某段期間有的現象，若繼續觀察，通常遲早該現象便中斷了，顯示並未存在一因果關係。但“莫里斯命中率高於 5 成”，與“快艇隊獲勝”卻的確可能相關。畢竟一球員有特別好的表現,是可能帶動士氣,增加球隊獲勝之機率。“有趣事件”的存在，仰賴的常是觀察入微。

在“愛麗絲鏡中奇遇”(Through the Looking-Glass，1871，路易斯卡羅(Lewis Carroll(1832−1898)著，中譯本很多)的第五章：

> Alice laughed. ‘There’s no use trying’, she said. ‘One can’t believe impossible things.’

> 'I daresay you haven't had much practice,' said the (White) Queen. 'When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast.'

早餐前能發現 6 件不可能的事件？令人嘆為觀止。人們常說資料探勘，如果能像白王后一般，很細心敏銳地觀察，自然可找出很多有趣的關係、現象，或屬於“不可能”的事件。只是所發現的關係或現象等，當然要加以確認，方能斷言此關係是穩定存在，或僅是偶發性的。

前面指出，相關不表就有因果關係。例如，有人注意到，當可樂消售量增大時，不少醫院腸胃科的門診人數，亦顯著增加。經統計分析後，也果真得到，可樂消售量與腸胃科之門診人數，二量相關性很高。那是喝可樂對腸胃不好嗎？不能驟下結論。比較可能的原因是，可樂銷量大，往往是天氣炎熱時。天氣炎熱，使食物易腐敗，導致較會吃壞肚子。氣溫便是所謂“干擾因子”(confounding factor)。當兩量相關性高時，須謹慎搜尋是否有干擾因子。另外，阿茲海默症(Alzheimer's disease)，乃一很令人擔心罹患的疾病，尤其是婦女。男性不必擔心嗎？事實上，是有研究指出，病患中約有 2/3 為女性。但這項研究的推論，應是“性別與罹患阿茲海默症的相關性很高”，不能輕率解讀成女性較易罹病。要知阿茲海默症，通常發生在年紀較大時，而女性平均壽命比男性長，因此罹病者以女性居多，這是合理的。在此壽命，

便是一干擾因子。有些相關並非假消息，只不過是無用的訊息。例如，曾有報導，“O 型、射手座、已婚男 最易中彩券”。前述 3 項條件都符合的，是否該趕緊去買彩券？可能不需要。單看血型一項，就知此項相關性沒啥用途，因台灣的居民裡，O 型血佔最多，超過 4 成，因此中獎者以 O 型血最多，只是合理而已。至於中獎者中，12 個星座各自佔的比例，很可能差異並不太大，但總有個星座比例最高，如今“剛好”是射手座而已。至於中獎者中，已婚男最多也沒什麼，很可能就是購買彩券者中，以已婚的男性佔最多。

“英國醫學雜誌”(British Medical Journal，縮寫 BMJ)，曾刊登一篇瑞典烏普薩拉大學(Uppsala University)的麥可森(Karl Michaelsson)教授，所領導完成的研究報告。此研究於 1987 至 1990 年間，針對 61,400 位婦女，記錄她們的飲食習慣，之後進行長達 20 年的追蹤觀察。該篇論文中指出，婦女若一天飲用 3 杯(約共 680cc)以上的牛奶，則死亡率比每天喝不到 1 杯的人高出 1 倍，且其中骨折的比率亦較高。人們一直以來，都以為喝牛奶是有益健康的，如今女性是否不該多喝牛奶了？別急，這仍要進一步研究才能下定論。因有些專家並不認同麥可森之觀測結果，他們認為很可能是婦女若有骨質疏鬆，常會被鼓勵多喝牛奶。但牛奶畢竟不是仙丹玉露，喝再多也免不了一死。結果死後卻把原因歸咎於牛奶，因果關係完全顛倒。是“骨質疏鬆”的人易骨折，因而易死亡。導致多喝牛奶與骨折及死亡率增加的關連性很強。至於牛奶喝多，是否會導致骨折增加？是否導致死亡率增加？這

才是一般人較關心的議題，卻非前述研究之結論。

身為研究人員，就得從事研究，並將成果發表。但成果都有用嗎？並不盡然。像前述那些花大功夫去研究，但成果的價值不太高之情況很多。再給一例。曾有人研究台灣自 1999 至 2008 年，1856 位 65 歲以上獨居老人的生活習性。發現經常逛街者，較不常逛街者，存活率高出 27%。此成果令人關切，難道“購物療法”對健康有益？如果真是這樣，獨居老人是否該多逛街以延年益壽？即使一身是病，是否該強忍病痛出門？實情說不定是對獨居的老人，身體狀況較佳者，才較方便自行逛街。而身體較好的人，存活率本來就較高。因果關係倒過來了。事實上，亦有人指出，獨居者只要樂於外出與人打交道，減少孤寂感，就有助健康了，逛街不過是這種方式之一。所以即使找到二量之相關性高，在運用前，也宜探討相關性產生之原因。

再看一例。曾有一則美國麻省理工學院(Massachusetts Institute of Technology)所做的研究之報導：單親媽媽撫養長大的男孩比較難有成就，日後也較容易離婚，製造下一代的單親家庭。離婚媽媽若有兒子，看了此報導後得趕快再婚嗎？以免當年所做的抉擇，將禍及愛兒。而望子成龍的媽媽，是否不論如何遇人不淑，都該吞忍下來，萬不可離婚？這可不見得。要知孩子的成就，乃與遺傳及成長環境關係密切，前述兩種作法，很可能都不見得有用，且還可能讓孩子將來的發展更差。因果關係易誤用的情況著實不少。如若去分析死亡的獨居老人之生活習性，很可能發現不抽菸的比例

很高。那能得到不抽菸者，於上了年紀後，將易自己一人過活的結論嗎？一般而言，即使多年的菸槍，一旦上了年紀，基於“貪生怕死”的原因，戒菸的比例不低，與是否獨居關係並不大。

但有時相關性是可善用的。曾有美國某大型連鎖超市，經由探勘顧客的購物清單，發現星期五晚上，啤酒與嬰兒尿布同時出現在購物清單的比例很高。難道是使用尿布者，較愛喝啤酒嗎？顯然不是。真正的原因很可能是，若家中有嬰兒，當爸爸的周末便常得留家幫忙照料，為打發時間，會看球賽轉播之類的節目。於是買尿布時，順便買些啤酒，以便看電視時享用。或者反過來，去買啤酒時順便買尿布，才不會在家球賽看一半，被太太差遣去買尿布。了解此原委後，超市可設計較佳的貨品動線，以同時增加啤酒及尿布的銷售量。經由統計分析，找出那些變數間相關性較高，並善加利用。警方辦案也常如此，從大量的數據中，找到種種線索。但切記相關性較高的兩變數間，其間並不必然便有因果關係？須更進一步探討方能得知真相。

有時二變數的相關性並非太明顯。生物相互間有一“密度制約”(density dependence)的族群效應。給一例子來看，這取自“半個地球：探尋生物多樣性及其保存之道”(Half-Earth: Our Planet’s Fight for Life，2016，愛德華威爾森著(Edward O. Wilson，1929−)，金恒鑣與王益真譯(2017))一書。一般人想到狼幾乎都是不好的，從小時候聽到的童話：小紅帽、三隻小豬，及狼和七隻小山羊等，到上學後學到的

有關狼的成語：聲名狼藉、引狼入室、狼狽為奸，及豺狼當道等，相信對狼的描述，極少是正面的。牧羊者更痛恨狼，欲除之而後快。美國的黃石國家公園(Yellowstone National Park)，也曾儘量除狼。大功告成後，1995 年，卻從加拿大東部引進一群狼。為什麼？原來狼不是書生，並非百無一用。對協助樹木生長，狼能扮演很正面的角色。在狼銷聲匿跡後，赤鹿(elk)等草食動物，便過著幸福快樂的日子，不受抑制地大幅成長。可憐白楊樹(Balsam poplar)苗，便被啃得光禿禿的。一隻狼在一星期內，可吃掉一整隻赤鹿(elk)，而一隻赤鹿在一星期內，可啃掉一大片白楊樹苗。公園裡每星期少一隻赤鹿沒什麼，但每星期少一大片白楊樹可就觸目驚心了。當公園內有狼群時，赤鹿能啃掉的白楊樹苗就會減少，白楊樹苗也就欣欣向榮。一旦移除狼群，赤鹿便多起來了，白楊樹苗的成長也就快速下降。明白此關係後，黃石國家公園便維持適當的狼群。

類似上述狼、鹿、白楊樹關係的例子很多。印度的孫德爾本斯國家公園(Sundarbans National Park)，有世界最大的紅樹林。公園內老虎的角色，便如同黃石國家公園裡的狼，會捕食並減少花鹿、野豬及獼猴等族群，間接保護了紅樹林。亦即紅樹林裡有了老虎，並不必過於厭惡，因會增加不少動植物之數量，有助維護公園的生物多樣性(Biodiversity)。另外，在“狼圖騰”(姜戎著，2004)一書中，描述大陸的內蒙草原，狼群經人類持續捕殺後，數量大減。在生態缺乏高等獵食者下，獺子(旱獺內蒙亞種)及地鼠等便大量繁殖。草原

被這些低等齧齒類動物啃光後，嚴重沙漠化，大部分已不適合放牧了。

網站搜尋功能日益提升的今日，可快速找出較相關之事物，以供決策者參考。如由 Google 網站，在短期間內，擁入多人查詢有關流感的病情，及如何用藥等，便能推斷流感可能快爆發了。這種利用搜尋流感的資料，與流感流行有相關性來預測，有時比官方的預測流感，更及時且經濟。只是兩因素間之相關性高，並不表其間有因果關係，這點我們已一再強調了。

23 相關性之考題

底下來看兩道學測數學科裡涉及“相關性”之試題。第一道是 104 學年的多選題：

小明參加某次路跑 10 公里組的比賽，下表為小明手錶所記錄之各公里的完成時間、平均心率及步數：

	完成時間	平均心率	步數
第一公里	5：00	161	990
第二公里	4：50	162	1000
第三公里	4：50	165	1005
第四公里	4：55	162	995
第五公里	4：40	171	1015
第六公里	4：41	170	1005
第七公里	4：35	173	1050
第八公里	4：35	181	1050
第九公里	4：40	171	1050
第十公里	4：34	188	1100

在這 10 公里的比賽過程，請依上述數據，選出正確選項。

(1)由每公里的平均心率得知小明最高心率為 188。

(2)小明此次路跑，每步距離的平均小於 1 公尺。

(3)每公里完成時間和每公里平均心率的相關係數為正相關。

(4)每公里步數和每公里平均心率的相關係數為正相關。

(5)每公里完成時間和每公里步數的相關係數為負相關。

“大考中心”提供的答案是(2)、(4)、(5)。

先檢視 5 個選項的敘述。由於題目中的數據，都是關於小明在“某次”路跑賽裡的資料，所以也只能得到有關小明在該次路跑的推論。但選項(1)及(2)，是問小明如何，選項(3)、(4)及(5)，卻皆未提到“小明”，兩相對照，會讓人以為(3)、(4)及(5)，是針對一般人的體能提問，這是命題者之疏失。另外，在選項(2)裡，於“小明”之後，有“此次路跑”4 字，選項(1)、(3)、(4)及(5)裡則沒有，再度，會讓人以為選項(1)、(3)、(4)及(5)，是針對一般情況之提問，而不僅是有關小明此次路跑的這組樣本。所以，若依現有題目之敘述，有學生遵循邏輯，未選(1)、(3)、(4)及(5)，應該算是對的。所以 5 個選項裡，都該有“小明此次路跑”幾字才完整。雖是屬於注重邏輯的數學科，但命題者對題目之敘述，顯然相當不謹慎。

要知這只是一次路跑的數據，若小明再跑一次，或繼續

跑 10 公里，將可能得到完全迥異的數據。因而所得之樣本相關係數，連正負號說不定都會反過來。這有如假設題目裡說“投擲一銅板 10 次，得到 5 個正面”，則若問“銅板出現正面的機率為 0.5？”，便不該選，要選也是還諸如提問“可以 0.5 做為銅板出現正面機率之估計值嗎？”。也就是不論投擲多少次，得到的都只是銅板出現正面機率之估計值。那能辯解機率有不同的意義，這裡乃指主觀的解釋嗎？若是這樣，則有沒有選此選項，便都該給分了。事實上，依據此次實驗所計算出來的只是樣本之相關係數，可做為母體相關係數之估計值。至於母體之“每公里完成時間和每公里平均心率”，及“每公里步數”和“每公里平均心率”，是否為正相關或負相關，並無法由題目所給之小明跑 10 公里後的數據得知。

再給一文字方面的問題。兩個變數間，才有所謂正相關、負相關，或無相關可言。至於相關係數，不過是一個數字，可能為正、負或 0。因此在選項(3)、(4)及(5)裡，問“相關係數是否為正相關(或負相關)”並不恰當，宜問相關係數是否為正(或負)；或者 “的相關係數”5 字全刪除。猜想命題者並不熟悉“相關係數”此一題材。

在學測如此大型的考試，命題者對文字的陳述顯得過於隨意，尚非本題最關鍵的缺失。要知就算題目寫得不清不楚，早已身經百戰的台灣中學生，大致能猜出命題者的意思。假設某人想觀察自己體重的變化，遂每天量測並記錄。能否想到該注意些什麼？有的！須儘量在相同的情況下量

測。例如，每天皆在剛起床時量測，這樣才較能相比。即使如此，每天起床時間可能有差異，或有些日子前一晚因應酬吃喝較多，因而就算採取固定時間量測，恐怕也不敢宣稱，確實做到每天在相同的情況下記錄，但至少已儘量了。如今題目一開始便敘明，小明是參加比賽。而眾所皆知，比賽有競爭，跑者大抵會依自己體能去配速。甚至，人非汽車也非機器，連續跑 10 公里，豈能維持每 1 公里的狀況都相同？因此，少有以如題所述的方式，收集個人數據並做分析。若每天在差不多同一時間跑 1 公里，量測 3 項數據，連跑 10 天，再分析所得的數據，還較合理些。出這種考題，可說易教壞學生的統計概念。話說回來，僅以少少的 10 筆數據，便大做分析，也未免不像個統計分析。

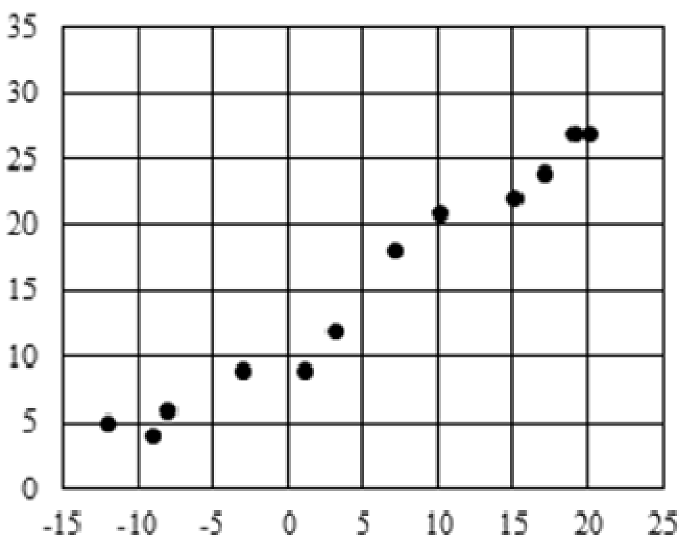
資料收集為統計分析裡，一重要的步驟。惟有秉持很嚴謹的態度，如此取得的數據，方能準確客觀，而得到的推論，也才較具參考價值。就如醫學上，一種新藥或新技術，想知其效果如何，便需找人做實驗。但並非徵求自願，來者不拒。實務上，不僅須謹慎挑選受測樣本，且實驗過程須有一定規範。而對某議題進行一項民調，也並非就站在商區街頭，任意找願意受訪者填寫問卷，或拿起電話便撥，誰接便問誰，若沒人接就再打下一通。取樣須很嚴謹，才能得到有意義的結果。

最後，對於上述考題，可否“假設”小明每一公里，都維持相同的狀態？亦即假設 10 筆數據(指完成時間、平均心率及步數)間為相互獨立。前面已說了，若加上這樣的假設，

則便是數學而非統計題目了。總之，考試畢竟引導學習，高中生若常接觸這類題目，將難具備統計素養。

再看一道是 106 學年學測的單選題：

下圖是某城市在 2016 年的各月最低溫(橫軸)與最高溫(縱軸)的散佈圖。



今以溫差(最高溫減最低溫)為橫軸且最高溫為縱軸重新繪製一散佈圖。試依此選出正確的選項。

(1)最高溫與溫差為正相關，且它們的相關性比最高

溫與最低溫的相關性強。

(2)最高溫與溫差為正相關，且它們的相關性比最高溫與最低溫的相關性弱。

(3)最高溫與溫差為負相關，且它們的相關性比最高溫與最低溫的相關性強。

(4)最高溫與溫差為負相關，且它們的相關性比最高溫與最低溫的相關性弱。

(5)最高溫與溫差為零相關。

“大考中心”提供的答案是(4)。

由所給最低溫 x 與最高溫 y 之散佈圖，不難看出(採目測近似值)最高溫 y 愈小時，溫差 y-x 愈大。如 y=5 時，x 為-12，因而 y-x= -17；y=27 時，x=19，20，y-x=8，7。由於是單選題，故知答案必為(3)或(4)之一，因最高溫與溫差只能為負相關。至於相關性強弱之比較，就不易以如此簡單的方式看出。新的圖就不給了，不過若繪出最高溫與溫差之散佈圖，將發現此新圖，12 個點大致由左上往右下散佈，即最高溫與溫差兩變數的確為負相關。另外，沿一直線散佈的情況，“感覺上”沒有原圖那麼明顯，因而判斷出最高溫與溫差的相關性，比最高溫與最低溫的相關性弱。

必須一提的是，題目所給散佈圖上的點畫得太大，在繪最高溫與溫差之散佈圖時，對緊張的考生不利就不說了。但

此題之缺失並不在此，而是沒什麼統計的味道。人們對父親身高與兒子身高、入學成績與大一成績、每日最高溫與翌日清晨最低能見度、每日心血管死亡人數與當日溫差等，會想了解其相關性。另外，求一天之最高溫與最低溫的相關性，也尚可想出理由。日溫差令人感興趣，月溫差可能就較無感了。至於為什麼會去求各月份最高溫與最低溫，此二極端值之相關性，本就已不清楚了，由此進而去求各月份最高溫與溫差之相關性，便更是目的不明了。顯然只是因有了 X 與 Y 之相關性，便去求 Y-X 與 Y 之相關性。再度，這是數學思維，而非統計思維。

近年統計被大量引進高中數學課程後，應是覺得高中學生該多懂些統計。只是由上述兩道學測題目，顯示在中學數學裡，統計不過被視為一類計算簡單的數學看待，如此是無法讓學生學到正確統計概念的。與其這樣，還不如將統計移出中學數學，免得讓學生學壞了。

附錄。

統計裡引進相關係數，主要是量測兩變數間之線性相關性(linear correlation)，包含強度及方向。相關係數取值介於 -1 至 1 間。若取正值，兩變數便稱正相關；若取負值，兩變數便稱負相關；而當相關係數為 0，則稱兩變數無相關。線性關係是一很簡單的關係，加上前述符合度量需求的性質，因此不只在統計裡，在很多科學的領域，都廣被採用來量測兩變數的線性相依程度。由於是為英國統計學家皮爾生(Karl pearson，1857−1936)首創，且為有別於統計裡其他不同定義的相關係數，有時稱為“皮爾生相關係數”(Pearson’s correlation coefficient，又稱 Pearson’s r)。

假設有兩個隨機變數 X，Y，其共變異數(covariance) $\mathrm{Cov}(X，Y)$，與兩變數的標準差 σ_X，σ_Y 之商，便是相關係數。在此

$$\mathrm{Cov}(X，Y)=E[(X-\mu_X)(Y-\mu_Y)]=E(XY)-\mu_X\mu_Y，$$

其中 μ_X 及 μ_Y，分別為 X，Y 之期望值。若以 ρ 表 X 與 Y 之相關係數，則

$$\rho=\mathrm{Cov}(X，Y)/(\sigma_X\sigma_Y)=E[(X-\mu_X)(Y-\mu_Y)/(\sigma_X\sigma_Y)$$

$$=E[((X-\mu_X)/\sigma_X)\cdot((Y-\mu_Y)/\sigma_Y)]。$$

由上式知，相關係數即兩隨機變數經“標準化”(減去期希望值，再除以標準差)後，乘積之期望值。由於標準化後之隨機變數，期望值為 0，標準差為 1，所以相關係數，可視為兩隨機變數標準化後之共變異數。兩非退化隨機變數 X 與 Y，相關係數要存在，先決條件是，X 與 Y 之期望值及變異數都存在。

可看出 $\mathrm{Cov}(X，X)=\mathrm{Var}(X)$。即自己跟自己的共變異數，就是變異數。因此共變異數的概念，乃變異數之推廣。如果變異數，是用來度量一隨機變數偏離期望值的程度，則共變異數，便能度量二隨機變數同時偏離各自期望值的程度。若

$X>\mu_X$時，較可能使 $Y>\mu_Y$，且若 $X<\mu_X$時，較可能使 $Y<\mu_Y$，也就是說 X 與 Y，有同時增大或同時減小之傾向，則$(X\text{-}\mu_X)(Y\text{-}\mu_Y)$便較可能是正的，因而它的期望值 Cov($X$，$Y$)也就較可能為正。反之，若較大的 X，有伴隨較小的 Y 之傾向，且較小的 X，有伴隨較大的 Y 之傾向，則$(X\text{-}\mu_X)(Y\text{-}\mu_Y)$便較可能是負的，因而它的期望值 Cov($X$，$Y$)，將較可能為負。也就是 Cov($X$，$Y$)之正負，能反映 X 與 Y 之增長方向，究竟傾向相同或相反。由於標準差為正值，所以相關係數與共變異數符號相同，且二者同時為 0，或同時不為 0。因此相關係數為正或為負，分別顯示 X 與 Y 之增大與減小的傾向，相同或相反。

我們說過，對二隨機變數 X 與 Y，共變異數是用來度量它們同時偏離各自期望值的程度。但與變異數類似，其值與所採用的尺度有關。例如，對於父與子身高的共變異數，身高單位採用公分，與採用公尺，前者之共變異數為後者之 1 萬$(=100^2)$倍。不過一旦除以兩標準差後，得到的相關係數，便無此困擾了。相關係數 ρ，永遠取值在區間[-1，1]。當 ρ 很接近 0，表 X 與 Y 的線性關係較弱；而當 ρ 很接近 1，或很接近-1，表 X 與 Y 有較強的線性關係，或者說相關性較強；ρ 較接近 0 時，便表 X 與 Y 有較弱的線性關係，或者說相關性較弱。事實上，當 $\rho=1$，或-1，X 與 Y 便有完美的線性關係，或者說 X 與 Y 為完全相關(completely correlated)。即對二隨機變數 X 與 Y，若 $\rho=1$，則存在常數 a，b，其中 $a>0$，使得 $Y=aX+b$；若 $\rho=-1$，則存在常數 a，b，其中 $a<0$，使得

$Y=aX+b$。曾有報導，同卵雙胞胎，若性別相同，則身高的相關係數很高，達到 0.95，這乃可以預期。至於大學成績與將來收入多寡的相關係數，就不見得太高了，且說不定是負的。

另外，當 X 與 Y 獨立時，$\mathrm{Cov}(X，Y)=0$，因而 $\rho(X，Y)=0$。但其逆不真。也就是說，共變異數(或相關係數)為 0，此時兩隨機變數為無相關，但不必然獨立。甚至，二隨機變數即使無相關，也不表二者沒有關係。例如，自區間[-1，1]隨機地取一個點，以 X 表之，再令 $Y=X^2$。Y 是 X 的平方，二者顯然關係無比密切，一旦知道 X，Y 便完全決定了。但底下來看 X 與 Y 卻為無相關。因 X 為一對稱的隨機變數，故 $E(X)=0$，且 $E(XY)=E(X^3)=0$，因而 $\mathrm{Cov}(X，Y)=0$，即得 X 與 Y 為無相關。這種例子很多，可參考一般機率論的書。仍要強調，相關係數雖說是量測二隨機變數之關連程度(degree of association)，但主要是反映二隨機變數間，線性關係之強度及正負，而非反映任何其他關係。

對於數據，亦可定義其相關係數。設有$(x_1，y_1)$，$(x_2，y_2)$，…，$(x_n，y_n)$，且分別以 $\bar{x}$，$\bar{y}$ 表其平均值，則數據 x's，與 y's 之相關係數，一般表示成 r，定義為

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^{1/2}[\sum_{i=1}^{n}(y_i-\bar{y})^2]^{1/2}}$$

例如，設有數據(1，5)，(3，9)，(4，7)，(5，1)，(7，13)，則 $\bar{x}=4$，$\bar{y}=7$，且

$$(1-4)^2+(3-4)^2+(4-4)^2+(5-4)^2+(7-4)^2=20，$$

$$(5-7)^2+(9-7)^2+(7-7)^2+(1-7)^2+(13-7)^2=80，$$

$$(1-4)(5-7)+(3-4)(9-7)+(4-4)(7-9)+(5-4)(1-7)+(7-4)(13-7)=16。$$

故得

$$r = 16/(20 \cdot 80)^{1/2}=0.4。$$

24 再談隨機取樣

生活裡，人們常提到“隨機取樣”一詞。如自區間[0,1]隨機取 1 點、自某班 n 位學生中，$n\geq 3$，隨機選 3 位學生去打掃、自 0，1，…，9 中，隨機取 6 個數字當提款卡密碼等。其中的隨機是什麼意思呢？隨機並非隨便，除非另有聲明，否則對於前述第一例，若以 X 表取中之點，則 X 有在區間[0,1]之均勻分佈；對於第二例，乃以“取出後不放回”的隨機取樣，抽取 3 位學生，即第 1 次，每位學生被取中的機率為 $1/n$，第 2 次，每位學生被取中的機率為 $1/(n\text{-}1)$，第 3 次，每位學生被取中的機率為 $1/(n\text{-}2)$，也就是對編號 1 至 n，每次取樣，會取中的號碼，皆有離散型的均勻分佈，分別是 $U[1,n]$，$U[1,n\text{-}1]$，$U[1,n\text{-}2]$，但各次取樣間並不相互獨立；對於第三例，則以“取出後放回”的隨機取樣，自 0，1，…，9，選取 6 碼。即每次取中的號碼，皆有離散型的均勻分佈 $U[0,9]$，且各次取樣間為相互獨立。“隨機取樣”可說常與均勻分佈相連結。

現代統計學的鼻祖費雪(Sir Ronald Aylmer Fisher，1890−1962)，曾提到下述故事。在 1920 年代後期，某日的下午茶時間，有位女士對一群科學家宣稱，奶茶的調製順序，對風味有很大的影響。把茶加進牛奶裡，與把牛奶加進茶裡，兩者喝起來口味大不相同。當時在座不乏各領域的泰斗，對這種說法莫不感到可笑。難道不知 $a+b=b+a$ 嗎？兩種

混合方式的化學成分，會有什麼差異？眾人皆醒我獨醉，費雪卻很當一回事地看待此女士的見解。他設計了一個實驗步驟，包括要準備多少杯奶茶，及該依照什麼順序給這位女士喝，以對這女士的說法做一檢定。這就是有名的“淑女品茶”(lady tasting tea)實驗，乃費雪在他引進“實驗設計”概念的著作裡，所舉之例子。世上奇人異事著實不少，還有人宣稱能以手指識字呢！撇開費雪，我們不妨想想，該如何設計一合理的程序，以檢定某人是否真能判斷“奶茶是先放奶或先放茶”？

首先，該準備幾杯奶茶來測試？這並無定論，不妨就先採 10 杯。其次，那幾杯先放奶，那幾杯先放茶？有人以為該隨機放，而既然是隨機放，則先放奶與先放茶的杯數便宜相同，也就是各 5 杯；至於各那 5 杯？就“隨機挑”，有人這樣提議。如何執行呢？將杯子以 1 至 10 編號，隨口唸出 5 個號碼，那些編號的杯子便先放奶，其餘當然便先放茶。對此方式，有人可能立即表示反對，覺得這樣不見得能符合隨機性，主張用抽籤。因 $C(10,5)=252$，這樣的組合數有 252 組，可準備 252 張字條，每張上有 1 至 10 中的某 5 個號碼，然後將字條全放進某袋子中。隨機抽取 1 張，其上 5 個號碼的杯子便先放奶，如此能全猜中的機率為 1/252。對此亦有人指出，做 252 張籤太麻煩了，可就做 1 至 10 等共 10 個號碼籤放進袋子中，攪和後，依序抽取 5 個籤，那 5 個編號的杯子便先放奶，其餘則先放茶。這的確簡易多了。只是又有人指出，這樣並不算太難猜，該每杯隨機先放奶或茶。如在

每杯奶茶製作前，先投擲 1 公正的銅板，若出現正面則先放奶，否則便先放茶。或者用一組亂數表，遇奇數則先放奶，偶數則先放茶。則全猜中的機率為 $1/2^{10}=1/1,024$。1,024 為 252 的 4.06 倍多，如此顯然難猜多了。若覺得 10 杯全中，約千分之 1 的機率仍不夠小，則可採 20 杯，且每杯均隨機先放奶或茶。由於 $2^{20}=1,048,576$，20 杯能全中的機率，比百萬分之 1 還小，真有人這麼神奇，恐怕就只好先相信他了。反正若是招搖撞騙，遲早會失手。當然也會有人以為，不見得須要求 20 杯都講對，畢竟人難免會犯錯，而一般犯點小錯是可以容忍的。至於能允許多大的犯錯機率？可事先設定一 α 值，α 為一不太大的正數，只要犯錯的機率不超過 α，便接受該女士“能分辨奶茶是先放奶或先放茶”之假設，並換算出至少要講對幾杯。上述的流程，看起來相當有邏輯，之後便發展出一套假設檢定(hypothesis testing)的理論。

要知人的天性，通常是沒有隨機性的，若僅憑腦海中“隨便”想到那個數字就說出來，雖自以為隨機，但所產生的數字，很可能是極不隨機的。有些中學教室，講台上放一籤筒，以供任課教師上課點學生回答問題時用。籤筒的使用，亦可分取出後放回及取出後不放回兩種。放籤筒是免得教師自以為每次都是隨機叫出一個號碼，學生卻發現，教師常就是點那幾個號碼。一般而言，諸如提款機的密碼等，隨機選號是最難猜中的。假設是 6 碼，有人以為誰會想到 1、2、3、4、5、6？遂以此做為密碼，偏偏人同此心。據統計，不僅在台灣，全世界最愛用的密碼即為 1、2、3、4、5、6。若

設定此為密碼，一旦提款卡遭竊，戶頭裡的錢，很快就被盜領了。

前面提過，隨機取樣常與均勻分佈相連結，因而口語裡的“均勻”，遂也屢會被聯想到隨機。有位母親在做芝麻餅，要讀大學的兒子幫忙灑芝麻，且要他灑得均勻些。兒子學過機率，想均勻就是隨機，遂隨機地灑芝麻。結果卻被母親責怪了，因有些地方芝麻很厚，有些地方很稀薄，相當不均勻。事實上，隨機的後果，常是不均勻。某新藥擬做實驗，負責人將受測者分成兩組，且宣稱採隨機分組。但若兩組分配的人數相同，將被懷疑並非真採隨機分組。教師點名學生，如果一學期下來，每位學生剛好都被點 1 次，也會被懷疑並非真的隨機點名。因隨機產生的號碼，即使是取出後放回，其中應很可能會有同號。底下給一例。

將 10 個球，隨機地投擲進 10 個箱子，則每箱中各有 1 球的機率相當小，為

$$10!/10^{10}=3,628,800/10^{10}=0.00036288，$$

即在隨機投擲下，極不容易很均勻地各箱中各恰有 1 球。現以 a 表上述機率，可求出事件“1 箱中有 3 球、7 箱中各有 1 球、2 空箱”之機率為 $60a$。此投擲後看起來相當“不均勻”的結果，發生之機率卻為很均勻的，每箱各有 1 球的機率之 60 倍。我們再給一些亦屬“不均勻”的事件之機率如下。

(a) 1 空箱的機率為 $45a$。

(b) 2 空箱的機率為 375*a*。

(c) 3 空箱的機率為 980*a*。

(d) 4 空箱的機率為(7609/8)*a*。

(e) 5 空箱的機率為(2,835/8)*a*。

(f) 6 空箱的機率為(6,821/144)*a*。

(g) 7 空箱的機率為(311/168)*a*。

(h) 8 空箱的機率為(73/5,760)*a*。

(i) 9 空箱的機率為 *a*/9!。

可看出一直到 7 個空箱，其發生的機率，都比每箱中各恰有 1 球的機率大。換個方式說，若自 10 個箱子中，每次隨機取 1 個，連取 10 次，且取出後放回，則 10 次中，取中的箱子集中在其中某 *b* 個，*b* 從 3 至 9，其機率都比 *b*=0(每箱都各被取中 1 次)容易發生，且集中在 7 個箱子(3 空箱)最容易，發生之機率為 980*a*。

類似的例子很普遍。如在著名的生日問題(birthday problem)裡指出，一團體裡，只要有 23 人以上，則其中至少有 2 人生日相同的機率，便大於 1/2。也就是假設將 23 個球，隨機地投擲進 365 個箱子(即忽略閏年的情況)，則有某箱中至少有 2 球，比球皆投進不同的箱中，更容易發生。雖然箱子多達 365 個，而僅投擲少少的 23 球，都進不同的箱子，

比至少有 2 球擲進同 1 箱中還難。事實止，一團體裡的 n 人生日皆相異之機率為

$$p_n=365\times364\times\cdots\times(365-n-1)/365^n。$$

因而至少有 2 人生日相同之機率為

$$1-p_n=1-(365\times364\times\cdots\times(365-n-1)/365^n)。$$

可解出 $n=22$ 時，$1-p_n\approx0.411$；$n=23$ 時，$1-p_n\approx0.507$；$n=40$ 時，$1-p_n\approx0.891$。由此知，一班只要有 23 人，有人生日相同的機率便超過 1/2 了，而若有 40 人以上，則至少有 2 人生日相同，根本輕而易舉，若發生完全不必驚訝，因機率高達約 0.891。如果班級大一點，如 $n=64$，則 $1-p_n\approx0.997$，幾乎必有 2 人生日相同了。

但若某人在小學時，班上才 20 多位同學，便有位生日與他相同，上國中後，尋尋覓覓，卻不只在班上，連全年級 200 個學生，都找不到一位生日與他相同，這是怎麼一回事？難道國中生的生日，較不隨機嗎？來看底下的推導。

假設一團體中，除某特定人外，另有 n 個人。則 n 個人中，至少有 1 人生日與該特定人相同之機率為

$$q_n=1-(364/365)^n。$$

令 $q_n\geq1/2$，解出 $n\geq253$。得知若欲 $q_n\geq1/2$，所需的人數，比我們想像的多很多。亦即在一團體中，有任 2 人生日相同容易，但對某特定人，尋找有人生日與他相同，可就難多了。

樂透彩的頭獎號碼是隨機產生，且屬於取出後不放回，其中也屢有令人訝異的結果產生。就以 42 取 6 的樂透彩為例。開出的 6 碼全為偶數之機率為

$$C(21,6)/C(42,6)=54,264/5,245,786\approx 0.0103，$$

並不算太大。但 6 碼全為奇數、6 碼全在 1 至 21 間，及 6 碼全在 22 至 42 間，機率也都約為 0.0103。再加上 6 碼全為 3 的倍數、6 碼全不為 3 的倍數，及某碼連續 5 期出現等，在開了很多期下，只要認真觀察，總會發現某些有趣現象，並不足為奇。除非經過統計檢定，否則不能就此斷言其中有弊，即號碼並非隨機產生。

在樂透彩裡，有些事件發生的可能性，遠比我們以為的大，仍以 42 取 6 的樂透彩為例。有連號之機率為

$$1-C(42-6+1,6)/C(42,6)=1-C(37,6)/C(42,6)\approx 0.5568>0.5。$$

即有連號比沒有連號更容易發生。因此不必因看到頭獎號碼裡，常出現連號，就產生懷疑。又，既然頭獎號碼裡，連號較可能比不連號多，那選號是否選連號較易中頭獎？這當然不對，任一組號碼中頭獎之機率皆為 1/C(42,6)。

樂透彩中 1 次頭獎就很難了，怎可能中 2 次？美國紐約時報於 1986 年 2 月 14 日，在頭版刊登一則大新聞，有位 Adams 女士，第二度獲紐澤西(New Jersey)州樂透彩頭獎。前一年(1985 年 10 月 24 日)她第一次中了 390 萬美元，第二次則獲 150 萬美元。兩次樂透彩，分別是 39 取 6，及 42 取 6，

中頭獎機率分別是

$$1/C(39,6)=1/3,262,623，$$

及

$$1/C(42,6)=1/5,245,786。$$

新聞中強調，任何人一生當中，能 2 次中頭獎的機率為 17 兆分之 1。這麼小的機率，顯然是由下述計算所得到：

$$1/3,262,623\times 1/5,245,786\approx 1/(17.115\times 10^{12})。$$

只是這樣對嗎？

若 Adams 一生中，就僅兩種樂透彩各買 1 張，則兩次皆中頭獎的機率，的確如上的約為 17 兆分之 1。但她其實每期都買好幾張，且買了好幾年。不必太多，假設對 39 取 6，及 42 取 6 的樂透彩，Adams 每星期分別買 3 張及 5 張。則每星期至少中 1 張頭獎的機率為

$$1-(1-3/3,262,623)(1-5/5,245,786)\approx 1.87265\times 10^{-6}。$$

這是大於百萬分之 1 的機率。現設 Adams 以上述方式連買 10 年。就忽略她在一星期內中 2 張頭獎的情況，因那機率實在太小了。則在 10 年的約 520 個星期裡，利用二項分佈，並取近似，得她 1 次頭獎皆未中的機率為

$$(1-1.87265\times 10^{-6})^{520}\approx \exp(-520\times 1.87265\times 10^{-6})\approx \exp(-9.737\times 10^{-4}),$$

恰中 1 次頭獎的機率約為

$$C(520,1)\times 1.87265\times 10^{-6}\times(1-1.87265\times 10^{-6})^{519}\approx 9.737\times 10^{-4}，$$

令 λ 表至少中 2 次頭獎的機率，則

$$\lambda\approx 1-\exp(-9.737\times 10^{-4})-9.737\times 10^{-4}\approx(1/2)(-9.737\times 10^{-4})^2$$

$$\approx 4.734\times 10^{-7}。$$

λ 約 1 千萬分之 4.734，此值當然是微乎其微。紐澤西州人口超過 8 百萬，假設有 k=50 萬(=5×10^5)人，皆以上述方式買樂透彩，又令 Y 表 10 年間至少 2 次中頭獎的人數，利用稀有事件法則，Y 之分佈可以 $P(\lambda k)$來近似。因

$$\lambda k=4.734\times 10^{-7}\times 5\times 10^5=0.2367，$$

故

$$P(Y\geq 1)=1-P(Y=0)=1-\exp(-0.2367)\approx 0.2108。$$

約 0.2108 的機率，已不能說太小了。同一人 2 次中頭獎，就算 10 年間未發生，20 年間就容易多了，若有 40 年，發生就完全不稀奇。何況要上紐約時報的頭版，並非只能是紐澤西州有人 2 次中頭獎，任何 1 州都行，而美國有多達 50 州呢！那為什麼並沒頻繁地看到有人 2 次中頭獎的報導？財不露白，大部分的人，應連中 1 次中頭獎都不想曝光吧！

總之，在隨機取樣下，不論放回或不放回，事件發生可能性之大小，直觀並非屢屢可靠。而在觀察次數夠大後，各種原

本罕見的事件，其發生常便稀鬆平常了。又，切記隨機取樣與“均勻”的連結，要很謹慎。

25 雙盲實驗

在有關疫苗的研發，不時會看到“解盲”與“雙盲試驗”(Double-blind study)。不是說資訊該透明，為人也該耳聰目明，盲有什麼好？怎會大喇喇地說一群人在盲目行事？

做決策不能沒有數據(data)，連小說中的大偵探福爾摩斯(Sherlock Holmes)，都曾說“Data! Data! Data! I can’t make bricks without clay。”在開發一新藥物的過程，不能閉門造車，須找人參與實驗，才能真正了解藥效及可能會有的各種副作用。而藥效如何，得經過比較，才知服藥與不服藥有何差別。免得發生如某些醫務人員所認為的，“感冒吃藥 1 周治癒，不吃藥 7 天自愈”。若服藥後之效果，不比不服藥之效果顯著地大；或服藥後，產生的各種副作用，如發燒、疲勞、肌肉痠痛、頭痛、腹瀉，及噁心嘔吐感等，比未服藥者顯著不佳，皆會令人對新藥提高警覺。因而受測者要分服用新藥的“實驗組”(treatment group，又稱處理組)，與不服用新藥的“對照組”(control group，又稱控制組或安慰劑組)。一般而言，醫護人員較傾向希望新藥有效，對服用新藥之受測者，會有意無意地更熱心照顧，此對受測者的健康，自然有很正面的影響；至於對照組，便有如被歸入放牛班，心理影響生理，於是健康狀態，遂難免會受到若干影響。較科學的方法，便會採雙盲試驗。即接受試驗的對象及研究人員，皆不知受測者中，那些屬於實驗組，那些屬於對照組。只有

屬於實驗組者才服用新藥，屬於對照組者就不用任何藥物嗎？不能這樣，這樣的話，誰屬於對照組就曝光了。對照組須注射(或服用)某中性物質(如生理食鹽水)，外觀與新藥相同，稱為安慰劑(placebo)。至於如何分組？宜將受測者隨機編入實驗組與對照組，而不宜讓受測者，採自願選組的方式。在所有資料都收集及分析完成後，研究人員才能獲知各受試者所屬組別，此即解盲(unblind)。藉由雙盲的設計，以減少整個實驗流程中，人為的干擾，並讓研究人員在分析數據時，能更加客觀。

“雙盲試驗”外，尚有單盲試驗(Single-blind study)及三盲試驗(Triple-blind study)。單盲只對接受試驗的對象保密。例如，在品酒時，如果看到酒瓶，知道所喝是某名酒，則評語通常不會差。因而一般會將瓶身矇住。此即單盲試驗，有一定的客觀性。缺點是研究人員(或工作人員)，基於主觀期望，言行舉止，有可能影響到參與試驗者之判斷。至於三盲試驗，就是除了雙盲，負責資料監測的人員，亦不知所有試驗對象中，誰屬實驗組，誰屬對照組。這當然比雙盲更嚴謹，且實驗結束後，會涉及兩次解盲。附帶一提，有些試驗並不易自始至終皆是雙盲。如若治療效果非常顯著，或治療的副作用非常明顯，則參與的實驗人員，便可能猜出受測者屬於那一組。在某些情況下，連受測者亦可能猜出自己屬於那一組。

雙盲試驗並不限用在判斷新藥之藥效。在刑事單位讓證人指認嫌犯之過程，通常為一對證人記憶的單盲測試。但為

免辦案者的一言一行，不自覺地對證人產生影響，有時在指認過程中，採雙盲測試。即參與辦案的人員，事先亦不知被指認的幾個人中，何者是嫌犯。總之，雙盲試驗是一種科學的決策方式，用途不小。

在一科學實驗中，如某教學方法是否有助提昇教學成效，便需對照組，僅有實驗組是不夠的。科學上的某些假設，欲進行實驗來驗證，也得有對照組，才易得到佐證。如欲顯示“吸菸會增加罹患肺癌的機率”之命題為真，若僅有實驗組，即使得到“某地區吸菸者罹患肺癌的可能性為 8%”，雖數據強烈支持此命題，但對吸菸者並不會產生太大的嚇阻作用。必須有對照組，得到諸如“在該地區不吸菸者罹患肺癌的可能性僅為 1%”，才能顯示“吸菸會增加罹患肺癌的機率”之命題為真。

我們說過取樣是很重要的，要儘量避免取樣偏差。對上述“吸菸會增加罹患肺癌的機率”之命題，如果實驗組中，有高比例是坐辦公室的白領階級；至於不吸菸的對照組中，卻有不小比例是在工廠的藍領階級，這便極不恰當。因坐辦公室總是工作環境較佳，至於對照組中的那些不吸菸者，說不定是因工作環境不佳，身體早就具各種大小毛病，在醫生規勸下遂戒菸，但罹患肺癌的可能性，自然仍不會太低。底下便來看一沙克疫苗實驗之例，以說明分組之重要。取材自羅夢娜(1987)(百農嘗一草—實驗設計之應用。科學月刊，第 18 卷第 5 期：338-340)一文。

1916 年，美國首次發現急性骨髓灰白質炎(poliomyelitis，簡稱 polio)的傳染病，由於患者一向以孩童居多，故此病又稱為小兒痲痺症。之後 40 年間，美國每年皆有幾萬人受到感染，且其中有幾千人因感染而死亡或身體部分變痲痺。例如，1952 年的大流行，美國全年感染者高達 5 萬 8 千人，其中有 3 千多人死亡，2 萬 1 千多人成終身殘疾。至 1950 年代時，已研發出數種對抗小兒痲痺的疫苗，其中又以沙克(Jonas Edward Salk，1914–1995)所研發的疫苗，看起來希望最濃。在實驗室裡，已證實它的安全性，並能產生對抗小兒痲痺的抗體。下一步便需要一臨床試驗來支持。1954 年，美國公共衛生局(United States Public Health Service)發起一極大規模的醫學實驗，約有 2 百萬孩童參與這實驗。其中約 50 萬孩童接受疫苗，約 100 萬孩童特地使其不接受疫苗，約 50 萬孩童則拒絕接受疫苗。從全美國較易受到小兒痲痺感染的一些學區中，選擇最易受到傳染的小學一、二、三年級(6-9 歲)的孩童來做實驗。接受疫苗的孩童便屬實驗組，未接受疫苗的孩童即屬於對照組。兩組人數雖不同，但沒關係，因所有比較，可基於人數比率。

如何分配孩童進入實驗組或對照組？由於孩童接受疫苗，需得到家長之同意，所以一種可能，是家長同意的孩童，便歸入實驗組，家長不同意的便歸入對照組。只是這樣的設計，看似簡單方便，卻可能產生取樣偏差。要知高收入家庭的家長，比低收入家庭的家長，通常更願意讓自己的子女接受疫苗進行實驗。但小兒痲痺是一種與衛生環境極相關的疾

病，低收入家庭的衛生條件難免較差，成長於這種環境的孩童，年幼時便可能具有由父母而來的抗體，因而有可能即感染過輕微的小兒麻痺，那些幸運的存活者，自此就免疫了；但高收入家庭，一般較注重衛生，來自這些家庭的孩童，較無機會在年幼時曾因感染過小兒麻痺，而產生抗體，於是日後感染得病的比率，便可能會較高些。為避免因諸如此類之外在因素，造成實驗結果的偏差，分組時，須盡量使兩組測試者，除注射的是疫苗或安慰劑之不同外，其他條件都愈相似愈好。如此才能在當實驗組的數據顯著比對照組更具優勢時，有信心地下結論說，差異純粹是因疫苗所造成。只是這些今日看起來很合理之分組方式，在早期可非普遍為科學家所熟知。

在 1954 年的實驗中，美國小兒麻痺基金會(National Foundation for Infantile Paralysis，簡稱 NFIP，乃成年後才罹患小兒麻痺的羅斯福總統，於他第二任任期中的 1938 年所創)，建議使用如下分組方式：將所有獲得家長同意的小學二年級孩童分到實驗組注射疫苗；至於一年級和三年級孩童則注射安慰劑，當作對照組。官大學問大，於是許多學區都採用 NFIP 的建議去分組。只是此分組方式存在兩大缺點：首先，小兒麻痺是一種傳染性疾病，可經由接觸傳染，因而每個年級染病比率差異往往不小。亦即有可能二年級孩童中的病例較多，這對疫苗效果便不利；反之，亦可能二年級孩童中的病例較少，這對疫苗效果便較有利。其次，如前所述，實驗組中之孩童，由於參與測試均獲得家長同意，因而家庭

因素可能高度影響實驗結果，使疫苗效果不易清楚浮現。

不少學區在得知 NFIP 分組方式之缺點後，便改採另一種分組方式，即所謂“雙盲隨機化對照實驗”(double blind randomized controlled experiment，簡稱 DBRC)。為了公平起見，對照組的孩童，須從與實驗組相同的群體中選出，也就是獲得家長同意的孩童群體。免得家庭因素的影響，會干擾疫苗效果。本來在分配孩童至實驗組或對照組時，為使二組孩童的各種條件愈相似愈好，最好也考慮到諸如家長收入、孩童之健康情形、性格、交友習慣等因素。但經驗顯示，在做這類選擇時，常常會產生人為判斷之偏差。所以最好的方法就是採用隨機化來分組。例如，採之前說過的，藉助亂數表，遇奇數則分入實驗組，遇偶數分入對照組。當然亦可經由投擲一公正的銅板來分組。隨機化對照實驗，是一種客觀的分組方式。再加上雙盲試驗，便構成“雙盲隨機化對照實驗”。除了可使選樣偏差降至最低外，還因其他因素都被控制住，因而能確定實驗結果的差異，是因兩組處理的方式不同(注射疫苗或安慰劑)所造成。

我們來比較兩種分組方式之實驗結果。依每 10 萬孩童之染病人數，對 NFIP 分組方式，實驗組(二年級)、對照組(一、三年級)及家長不同意組(二年級)，各為 25、54 及 44 人；對 DBRC 分組方式，實驗組、對照組及家長不同意組，各為 28、71 及 46 人。DBRC 的分組方式，疫苗效果顯著不少。

之後沙賓(Albert Bruce Sabin，1906−1993)研發出對抗小

兒麻痺症的沙賓疫苗(Sabin vaccine)。在科學家的齊心努力下，自 1979 年起，美國再無自然產生的小兒麻痺症病例。

時至今日，新藥研發過程需先申請人體臨床試驗階段(Investigational New Drug)，才能進行臨床試驗分期階段(Phase Ⅰ、Phase Ⅱ、Phase Ⅲ，及 Phase Ⅳ，共 4 期)，而分組方式皆採“雙盲隨機化對照實驗”。在第Ⅲ期臨床試驗(Phase Ⅲ)過程中，若證實藥物療效，便可申請新藥上市許可。在第Ⅰ期的臨床試驗，主要目的是監控藥物的安全性，藉由此階段的試驗，探討藥物對人體所起的生化及物理作用，並觀察人體對藥品作用的過程，其中包括如何吸收、分配及新陳代謝等。另外，藉著第Ⅰ期的觀察，研究者將找出適當的治療劑量及給藥時程。在第Ⅱ期的臨床試驗，乃在探討新藥或治療方法(或標準方法的新用法)之安全性，及有效性，並評估其對人體的影響。第 II 期研究的試驗通常把焦點放在特定的醫療狀況。此期試驗規模往往不太大。通常第 I 期及第 II 期的執行成果，顯示具有前景的試驗，才會進入第Ⅲ期的臨床試驗，在第Ⅲ期將更進一步評估新藥(或治療方法)的有效性及安全性，並比較新藥(或治療方法)與原有的標準療法，此階段試驗受測者人數會多些。第Ⅳ期的臨床試驗，乃進一步評估長期治療的安全性和有效性的試驗。通常在治療已獲准作為標準用法之後進行。第 IV 期試驗參與人數也可能不少。一般新藥上市前，均得經此 4 期嚴謹的臨床試驗。

26 交叉分析

投擲 1 公正銅板 n 次，則出現的正面數 X 之期望值 $\mu=n/2$，標準差 $\sigma=n^{1/2}/2$。如果 n=10,000，則 μ=5,000，σ=50。由中央極限定理，將約有 0.9544 的機率，X 介於[4,900，5,100]，此區間長度達 200。至於正面數出現的相對頻率 X/n，則亦有約 0.9544 的機率，介於[0.49，0.51]。此區間長度為 0.02，並沒想像中的短。而且，尚有約 0.0456 的機率，X 及 X/n，分別不落在前述二區間中。現令 n 增大些，如 n=100,000,000。則 μ=50,000,000，σ=5,000。仍由中央極限定理，將約有同樣 0.9544 的機率，X 介於[49,990,000，50,010,000]。即出現的正面數 X 落在一更大的範圍，寬達 20,000，即可偏離期望值 50,000,000 更遠。至於正面數出現的相對頻率 X/n，則有約 0.9544 的機率，介於[0.4999，0.5001]，長度僅 0.0002，即 X/n 將更接近期望值 1/2。可看出隨著投擲數 n 之不斷增大，出現的正面數 X，將可能愈來愈偏離期望值 $n/2$，但正面數出現的相對頻率 X/n，則將有高機率，愈來愈接近期望值 1/2。但不論 n 有多大，X/n 就是很難剛好等於 1/2。也就是即使有些偏差，仍是正常的，不會因而懷疑銅板的公正性。反倒是，若 X/n 過度接近 1/2，將可能被認為其中有弊。

曾有好幾年，高中數學裡有“信賴區間”的題材。要知即使在大學數學系的統計課程，信賴區間通常置於教科書的

後半部，有夠完整的鋪陳後，才進入信賴區間。而為了學習信賴區間，所引進的中央極限定理，對大學數學系的學生，更絕非容易的題材。既然如此，那為何此二題材皆會進入高中？可能有些課綱委員，認為高中生該多懂些統計，至於得學那些？則可能因見到媒體上常會報導各式各樣的民調結果，於是覺得信賴區間乃如國民統計，高中生皆該理解。雖立意良善，卻未曾考慮此題材是否適合高中生。於是在 95 課綱時，信賴區間連同中央極限定理，便堂而皇之地進入高中數學了。又應是見到執行民調時，常會同時比較各族群對同一議題之支持度，此即“交叉分析”(Cross Analysis)。於是高中的選修數學裡，便有了交叉分析。由於很快便發現高中生不適合學此題材，99 課綱便將交叉分析拿掉了。至於信賴區間，不少高中教師仍一直努力與其奮鬥，直到幾年前，在撰寫 108 課綱時，才徹底放棄此題材。至此讓高中師生困擾多年的信賴區間(及中央極限定理)，才不得不離開高中數學了。以昭炯戒，雖一切都已成為過去，大家仍不妨思索，是什麼原因，使得當初信賴區間、交叉分析及中央極限定理，幾個那麼深的題材，會被放進高中數學？

交叉分析為何不適合在高中？底下為某教科書中的例子，而這在當時各版教科書中，是很典型的例子。先給出某校入學考試的所謂列聯表(Contingency table，又稱交叉表(cross tabulation))：

	錄取人數(A)	未錄取人數(F)	合計
男生(B)	24	36	60
女生(G)	36	54	90
合計	60	90	150

由此得

男生錄取率為 $P(\mathrm{A}|\mathrm{B})=40\%$，

女生錄取率為 $P(\mathrm{A}|\mathrm{G})=40\%$。

然後就說男生女生錄取率沒有差異。有些教科書會加上類如底下的一句："至於比例不相等時，是否就代表男女生錄取率有差異，留待日後再學習。"這樣的寫法，顯示既具學術良知，卻又無可奈何。要知大學的犯罪防治學系，在有關嫌犯辨識的課程，如果所舉的例子，都是"符合 xx 條件者，便無犯罪之嫌"，且說"若不符合 xx 條件者，是否就有犯罪之嫌，留待日後再學習"，則這門課豈有何大用？像這種既講不清楚，也無法讓學生學到任何正確統計概念的交叉分析，居然能在眾多數學專家盯著下，溜進高中數學，實在令人難以理解。

這其中有兩點必須指出。首先，怎可由男女錄取率的"相等與否"，來判定錄取與男女性別是否有關？這毫無統計思維。我們已多次強調。如果投擲銅板 100 次，恰好出現 50 個正面及 50 個反面，並不會讓人相信銅板為公正，反而較易讓人懷疑其中有作假。像教科書這種例子看多後，恐會

使初學者誤以為，事件出現次數的相對頻率，就是該等於事件的機率。要知，除非事先設定男女錄取率一定要相同(這時男女的“錄取標準”，就很難相同了)，否則即使用抽籤(這時錄取與否總該跟性別無關了)，來決定錄取名單，都不能保證抽出的男女錄取率相同。更不要說，追求男女平等，應是追求男女“錄取標準”相同，而非“錄取率”相同。其次是較微妙的一點。這只是一次錄取的數據，不宜過度引申。即一次考試的錄取率，豈能使用條件機率的符號 $P(A|B)$，及 $P(A|G)$？來看個例子。投擲一銅板 100 次，出現 52 次正面，可將 52/100=0.52 當做銅板正面出現機率之估計值，但不會理所當然地視此值為銅板正面出現之機率。這點人們平常大都能了解，像是不能將一次民調的支持率，當做候選人的得票率。但不知何以統計只要一擺進高中數學課程中，人們往往就連常識都失去了。

我們再給一有名的例子。交叉分析並不只能用來檢定各類之比率相同與否，如樂透彩開了多期後，檢定 1 至 42，42 個號碼出現之頻率是否相同。也可用來檢定各類事件(如連號)之出現，是否符合該有之比率，用途廣泛。著名的遺傳學家孟德爾(Gregor Johann Mondel，1822−1884)，有一關於豌豆生長的實驗。他將圓黃(round yellow)種子的豌豆，與縐綠(wrinkled green)種子的豌豆雜交。依其理論，會生長出圓黃、圓綠、縐黃及縐綠種子的後代之比率，應分別為

9/16=56.25%，3/16=18.75%，3/16=18.75%及 1/16=6.25%。

經由一組有 556 個樣本的實驗，他得到如下表中的後代之觀測比率與預期比率。

	圓黃	圓綠	縐黃	縐綠	合計
後代數	315	108	101	32	556
觀測比率	56.65%	19.42%	18.17%	5.76%	100%
預期比率	56.25%	18.75%	18.75%	6.25%	100%

乍看之下，4 種豌豆觀測到的後代比率，與預期比率都有些差異，並不吻合該有之比率。但經過所謂“卡方檢定”(Chi-squared test，這當然無法在高中數學裡講授)，即使可忍受的誤差 α 值大到 0.90(一般 α 取為 0.001、0.005、0.01，或 0.05 等，很少有大於 0.1 者)，都無法拒絕

虛無假設：孟德爾的理論為正確。

舉例來看。投擲一銅板 10,000 次，若出現的正面數落在區間 [4,994，5,006]，便是即使將誤差 α 取為 0.90，都無法拒絕虛無假設：銅板為公正。正面數離 5,000 那麼近，不但不會讓人相信銅板公正，反而可能立即被懷疑根本沒投擲，數據乃造假。同樣，孟德爾也踢到鐵板。由於此實驗結果與預期太吻合(fit too well，上述高中數學教科書，所舉有關錄取率的例子，則是完美吻合)，曾引起著名統計學者費雪的懷疑，認為孟德爾可能是持續重覆做實驗，直到結果看起來“很好”才停止，然後只公佈結果“最好”的那組數據。偷雞不成蝕把米，這就是我們已多次強調的，對於隨機實驗，若結果與理論值過於一致，反而會讓人懷疑其中有弊。而眾所皆知，

不論上述豌豆雜交或投擲銅板的實驗，當被質疑時，若重做一次實驗，想再度得到那麼“漂亮”的數據，機率可是微乎其微。

27 假設檢定

數學裡常在證明“若 p 則 q”形式的命題，一旦證出，便毫無例外地成立。除非指出證明中有錯誤，否則企圖推翻該命題，將徒勞而無功。歐幾里得(Euclid，約西元前 325–265 年)將在他之前各數學家所證明出之命題，以及若干他自己所推導出之結果，彙集成“幾何原本”(The Elements)一書，內容主要是幾何，也包含一些代數及數論的結果。此書在今日仍相當有價值。所以至少在兩千多年前，人們就懂得數學證明了。隨機世界裡，卻很少在證明。某銅板是否公正，即使投擲再多次也都無法確定，最後有可能是基於沒什麼理由可拒絕，而接受是公正；某開發出的新冠疫苗是否有效？即使通過檢驗，也上市了，可能仍有不少人不信其藥效；某殺人嫌犯是否有罪？就算最後被判定無罪釋放，恐怕還是有人堅信他有罪。在隨機世界裡，無法證明的事，俯拾皆是。只能想成處處是假設，就看接受那一個。

既然處處是假設，那如何決定是否接受某假設？可採“無罪推定原則”！這是今日世界各國法庭普遍採取的原則。我國刑事訴訟法第 154 條，“被告未經審判證明有罪確定前，推定其為無罪。犯罪事實應依證據認定之，無證據不得認定犯罪事實。”這就是“無罪推定原則”。台灣並沒那麼先進，此原則出現在台灣法庭的歷史並未太久。我國刑事訴訟法，直到 2001 年，才廢止原於民國 25 年，所立下的有

罪推定原則。自此被起訴之嫌犯，即使千夫所指，法官仍會先假設他無罪。在無罪的假設下，何以有這些與嫌犯有關之不尋常的事件(或說顯著事件(significant event)，即發生機率很小的事件)發生？嫌犯若無法交待清楚，則被判有罪便怨不了別人了。顯著事件若發生，會引人注意，至於尋常的事件，即發生機率不算太小的事件，其發生自然不會令人太在意。要注意的是顯著或不顯著，乃依發生機率而定，與“值”的大小無關。如賽跑成績進步 0.1 秒，算顯著嗎？這並無法回答。對於男子 100 公尺短跑的世界紀錄，1968 年為 9.95 秒，目前則為 2009 年所創之 9.58 秒。每 10 年平均進步不到 0.1 秒。甚至自 2009 年起，紀錄一直高懸，因此若在某次比賽中，有選手將紀錄推進 0.05 秒，雖有如一剎那的 0.05 秒，不會讓人有何感覺，卻會被視為顯著進步。但 0.05 秒，對馬拉松(marathon，路程為 41.195 公里)或半程馬拉松(路程為 21.0975 公里)賽，便微不足道了。2021 年 8 月 29 日有則報導，在北愛爾蘭安特里姆海岸(Antrim Coast)賽事中，剛滿 22 歲，來自衣索比亞(Ethiopia)的耶華勞(Yalemzerf Yehualaw)，打破女子半程馬拉松(halfmarathon，路程為 21.0975 公里)賽之世界紀錄。耶華勞以 63 分 43 秒的成績，比今年初於伊斯坦堡(Istanbul)半程馬拉松賽裡，肯亞(Kenya)選手切普恩傑蒂奇(Ruth Chepngetich)，所創的世界紀錄 64 分 02 秒快了 19 秒。

統計裡假設檢定的想法，就是在無罪推定的原則下，如果某顯著事件發生，那原本的假設，就可放棄了。至於多小

的發生機率才算顯著？要事先設定，而這又視情況而定。若涉及食品添加物是否超過含量，則 5%的發生機率通常便不算小了，因較該保護的是消費者而非廠商；若是關於死刑的判決，則連百萬分之 1 的誤判機率可能都覺得太大，畢竟人命關天。執行假設檢定時會不會誤判？當然會，連法官也會誤判。法官有什麼樣的誤判？有兩類。第一類是嫌犯無罪卻被判有罪，第二類是嫌犯有罪卻被判無罪。民主時代，相當注重人權，通常第一類誤判被視為較嚴重。理論上兩類誤判都不應該發生，都須儘量減少發生的可能性，但通常就是很難將兩類誤判的機率同時降低。想想如果過度把關第一類誤判，則實際有罪，卻由於證據不夠強，因而被縱放的，將大幅度增加，這並非好事；反之，若過度把關第二類誤判，則將導致寧可錯殺 1 千，不可誤放 1 人的後果，製造出一個又一個的冤屈。宋朝歐陽修說他父親當年為官時，對已被判死囚的犯人，會反覆審閱其案件，為的是"求其生而不得，則死者與我皆無恨也。"無罪推定正是秉持盡全力為死囚犯求其生的精神，即使這樣都還難免誤判了，怎可先射箭再畫靶，即一開始便認定嫌犯該死，然後檢驗證據是否吻合？

統計學裡，假設檢定的理論與架構，是波蘭裔的統計學家奈曼(Jerzy Neyman，1894–1981)，及英國統計學家皮爾生(Egon S. Pearson，1895–1980，為 Karl Pearson 之子)，於 1933 年，給出著名的奈曼–皮爾生引理(Neyman-Pearson lemma)所奠定的。他們設計了一套檢定的流程。一開始先確定二假設，即虛無假設(null hypothesis)，以 H_0 表之，與對立假設

(alternative hypothesis)，以 H_a 表之。虛無假設通常表現況，或傾向推翻的；而對立假設則表傾向接受的。雖想推翻虛無假設、明明已對虛無假設充滿懷疑，卻儘量保護，不讓它輕易被推翻。如此一旦推翻，才能減少不服。統計學裡的假設檢定，乃依所觀測到的樣本(即數據)，來決定究竟要接受虛無假設或對立假設。但如法官判案，豈能永不犯錯？若虛無假設為真，卻接受對立假設，稱為第一型錯誤(type I error)；若對立假設為真，卻接受虛無假設，稱為第二型錯誤(type II error)。先設定一能容忍的第一型錯誤之機率，通常以 α 表之，稱為顯著水準(significance level)。常取的 α 值為 0.01，0.05，或 0.1 等，當然若要取其他 α 值也沒有不行。給定 α 後，便要決定拒絕域(rejection region，或 critical region)，即決定何時拒絕 H_0 而接受 H_a。

在 H_0 為真的假設下，出現的結果，若落在拒絕域，即稱得到顯著的結果。小機率事件發生，豈可等閒視之？亦即結果若顯著，便該拒絕 H_0。出現的結果，若沒有落在拒絕域，即不夠顯著，表得到發生機率不算太小的結果，一尋常事件，尚不足以撼動 H_0 的假設，遂仍接受 H_0。以產品檢驗為例。若產品明明有高比率不合格，取樣時卻大都取中合格產品，廠商豈不就僥倖過關？是這樣沒錯，隨機世界的確充滿機運。但怎會次次好運？若廠商不思改進，則總有通不過檢驗的時候，那時品牌信譽便受損了。

就好像拿到相同的數據，不同的人可有不同的推論。對同一 α，拒絕域的選取，並不唯一。若能有使第二型錯誤的

機率值 β 最小的拒絕域，當然相當完美。此拒絕域，稱為顯著水準不超過 α 下之"最佳拒絕域"(best critical region)。在某些條件下，統計學裡有一套找到最佳拒絕域的方法。那些以為假設檢定的主題很深奧者，往往是對找最佳拒絕域心有餘悸。不過在很多情況下，憑直觀所決定的拒絕域，常便是最佳拒絕域。更何況人生有多少機會在找最佳拒絕域呢？有如女生決定對象，常不過就是從兩、三個已向她告白的男子中挑選一個，什麼時候會尋找在全世界裡的最佳者？

至於虛無假設命名的由來，乃是因那根本是一空的假設。試想如果產品其實是合格的，主管的政府單位，卻懷疑其成分有問題，非要抽樣來檢驗，大費周章後，卻宣佈該產品合格(接受 H_0)，廠商不罵擾民、損害商譽才怪；又如檢察官大都在意自己起訴案件定罪率之高低，如果被他起訴者，最後卻被法庭宣判無罪(接受 H_0)，檢察官將灰頭土臉。接受虛無假設，往往表示白忙一場，天下本無事，庸人自擾之。虛無假設是執行檢定者，一點都不想接受的假設。

假設檢定裡，通常是先給定一 α，然後由所得觀測值(數據)，做出接受或拒絕 H_0 之推論。但有時對這樣的展示推論，被認為不夠詳實，因若接受 H_0，並不知是勉強地接受，或信心滿滿地接受。這便產生了 p-值(p-value)的概念。所謂 p-值，乃在 H_0 之為真之下，會得到比觀測值，至少同樣極端的數據之機率。求出 p-值後，不同的決策者，可依其所設定的 α 值，而決定是否接受 H_0。

底下來展示一憑直觀，找到最佳拒絕域之例。考慮檢定銅板出現正面的機率 p。設 H_0：$p=1/2$，H_a：$p\neq 1/2$。即擬檢定此是否為一公正銅板。持續投擲銅板 n 次，以 X 表所得正面數，則 X 有 $B(n，p)$分佈。當 H_0 為真，即 $p=1/2$，則觀測到的 X，較可能落在期望值 $n/2$ 的附近。所以直觀上，當 X 偏離 $n/2$ 較大時，便該拒絕 H_0。由是取拒絕域為

$$\{|X-n/2|\geq c\}=\{X\geq n/2+c，\text{或 } X\leq n/2-c\}，$$

其中 c 將由 n 及 α 來決定。現設 $n=100$，且取 $\alpha=0.05$。再度，因 n 較大時，二項分佈的機率值就不太好算了，以常態分佈來近似，得 c 約為 9.8，取整數 $c=10$。如此拒絕域$=\{X\geq 60$，或 $X\leq 40\}$，再回頭求出實際的 α 值約為 0.0456。對離散型分佈，有時無法取到剛好能達到所給 α 值之拒絕域。前述拒絕域，便為顯著水準不超過 0.05 下之最佳拒絕域。設觀測到 $X=61$，則 p-值=0.0278，此時若 α 取得比 0.0278 小，如 0.01，便無法拒絕 H_0 了；若觀測到 $X=65$，則 p-值=0.0026，相當小，此時除非 α 取得比 0.0026 小，否則都無法拒絕 H_0 了。另外，若取 $\alpha=0.01$，則 c 約為 12.88，所以取 $c=13$。如此拒絕域$=\{X\geq 63$，或 $X\leq 37\}$，此時實際的 α 值約為 0.0094。此為顯著水準不超過 0.01 下之最佳拒絕域。在同樣的 n 之下，所取之 α 值愈小，表 H_0 愈被保護，因此拒絕域將愈小，即愈不容易拒絕 H_0。當 $\alpha=0.01$，如果投擲銅板 100 次，得到 62 個正面，比在 H_0 下(銅板為公正)的期望值 50 多了 12，即超過 24%，感覺上很偏差，卻仍得接受此銅板為公正。沒辦法，那是因 α 取得太小之故。解決之道是加大 n。假設取 $n=10,000$，則

在 H_0 為真之下，X 之期望值=5,000。當 α=0.01 時，c 約為 128.8，故取 c=129，且拒絕域=$\{X\geq 5{,}129$，或 $X\leq 4{,}871\}$。此時正面數 X，只要比 5,000 偏離逾 129/5,000=2.58%，就得拒絕 H_0 了。至於有關求第二型錯誤的機率，比較複雜些，在此就不討論了。

無罪推定原則倒也非處處通行無阻。例如，由於學位論文不時發生抄襲事件，今日不少大學校要求學生畢業前，必須提交自己的論文進行檢測，以確定是否有抄襲行為。有些學生則抗議此作法，因這意味著有罪推定。除此之外，大部分的科學領域，於作決策時，大抵都採無罪推定原則。科學家不時宣佈一些新發現，如接觸某殺蟲劑會使人罹患帕金森氏症的機率增加，及過胖的中年人罹患失智症的風險較低等，其中宣佈的增加或較低，其依據可說就是執行一項假設檢定後之推論。即使生活裡，人們在面臨該接受那一選項時，往往也是憑藉假設檢定。假設檢定此科學性的思維，乃一適合讓國民提早學習的統計方法。

28 再探假設檢定

機率與統計有何不同？考卷上有道題，“設有一公正銅板，獨立地投擲 20 次，試求至少出現 17 個正面的機率。”不必在乎是否真有一個公正銅板，也不必實際去投擲，學生很快便能算出正確答案 $1{,}351/2^{20}$。此為機率問題，在給定的前提(銅板為公正)下，去推導出指定事件的機率。其中前提是給的，不必去懷疑是否真有此前提，這與數學裡一向的推導，其精神是一致的。再看一情況。若有人獨立地投擲某銅板 20 次，由於事先看不出銅板有何特別，因而合理地視為公正。結果一擲之下，得到 17 個正面。因正面數過多，遂懷疑該銅板並非公正，出現正面的機率可能大於 0.5，於是去執行一個假設檢定。這便是統計問題，見可疑追查到底，由觀測到的結果，去檢驗前提(銅板為公正)是否可接受。

對於銅板，若投擲出較偏差的結果，便懷疑其公正性；若投擲出的正面數，約為一半的投擲數，便大致相信其公正性。因正面數出現的多寡，乃由銅板出現正面的機率而定，且銅板會無怨無悔地讓人投擲，並維持同一出現正面的機率。至於投擲的人會不會搞鬼？除非有證據，否則便不考慮此因素，當然也是認為銅板的投擲，想搞鬼沒那麼容易。但銅板之外的假設檢定，就得很謹慎地執行。如曾有一則報導，標題是“德國科學家證實，喝綠茶可以減肥”。看到這種新聞，立即會令人想到假設檢定，要找人做實驗。但人可

不是銅板，每個人之體質各不相同、人也不見得會安份地聽命行事，且人除了綠茶外還會吃喝很多食物，這些都使減肥若要與喝綠茶連結，實驗必須嚴謹地設計。

我們以下例來說明，假設檢定的推論，應用時須謹慎。A 君某日心血來潮買了 1 張 49 取 6 的樂透彩，而居然就中了頭獎。頭獎要 6 碼全吻合，不計順序。因此中頭獎的機率為

$$1/C(49，6)=1/13,983,816。$$

將近 1 千 4 百萬分之 1 的機率，卻僅買 1 張就中了。A 君的好友 B 君，他每期都買好幾張，幾年下來，卻只曾中過幾個小獎。B 君對 A 君說，“你一定是買通樂透彩公司的員工，或能算牌之類的，否則豈可能一買就中？”A 君當然否認。兩人都學過統計，B 君說“我們來做一假設檢定”。A 君心裡坦蕩蕩，欣然同意，但強調“虛無假設須取成我是清白的”。B 君毫不猶豫地說“那當然。”虛無假設是被保護的，A 君遂安心地等著看 B 君“證實”他清白。現令

H_0：A 君沒作弊，H_a：A 君作弊，

分別表虛無假設及對立假設。此處作弊的意思很廣，借助任何外力(包含作假)都算。至於拒絕域要取成什麼？要知只要是合理的拒絕域，便該包含 A 君中頭獎，這是一明顯可能觀測到的結果。如今在 H_0 為真下，觀測到 A 君中頭獎，此機率即

$$p\text{-值}=1/13,983,816。$$

B 君認為實務上顯著水準 α 極少有設定這麼小的，所以在任一合理的 α 之下，皆該拒絕 H_0，而接受 H_a。居然被證實作弊！怎會這樣？A 君愣住了。

事實上，眾所皆知，樂透彩只要銷售量夠大，則“有人”中頭獎，幾乎可說是必然。如今這個“有人”，不過恰好就是 A 君而已。投擲銅板出現正面數之多寡，與銅板出現正面的機率，極度相關。但樂透彩要中頭獎，並非只能靠作弊，還有運氣可依賴。所以就算 B 君如上得到 A 君作弊的推論，大部分的人，恐怕仍是相信 A 君就只是好運而已。否則用相同的方法，將得到歷來每位樂透彩的頭獎得主，都是作弊之推論。但若下一期，A 君仍僅買 1 張，且又中頭獎，當然便會被強烈懷疑其中有弊。第一次中頭獎，A 君僅會被視為不過是“有人”好運臨門的那位。而前面已指出，一旦銷售量夠大，“有人”好運中頭獎，乃毫不稀奇。但 A 君第二次中頭獎，人們的認知將是，有一“特定的人”中頭獎，其機率為 1/13,983,816。運氣好得驚人，被調查就沒什麼好覺得委曲的。只是若無其他佐證，也不能就因 A 君運氣過度好，便判定他作弊。其實對前述投擲銅板之例，即使投擲 20 次出現 20 個正面，雖對大部分實務中的顯著水準 α(只要大於 $1/1{,}048{,}576(=1/2^{20})$)，都將拒絕“銅板為公正”。但銅板是否真的不公正？並不得而知，只能說在此結果(出現 20 個正面)下，接受銅板不為公正，是合理的。假設檢定並無法證明那一假設為真。這是何以科學家一旦接受某假設，習於宣稱“證實”(而不說證明)某假設成立，如前述“證實喝綠茶可

以減肥”。

如同遺傳學家孟德爾之豌豆生長實驗，由於結果與預期太吻合，遂被懷疑可能是持續重覆做實驗，直到結果看起來很好才停止，然後只公佈結果較好的那組數據。假設檢定也不能只公佈想要的數據。否則若顯著水準 α 取為 0.05，則即使 H_0 為真，可預期平均做 20 次實驗，便將有 1 次能拒絕 H_0；就算 α 取為 0.01，且 H_0 為真，仍平均做 100 次實驗，便將有 1 次能拒絕 H_0。舉個例來看。假設某公司研發出一種新飲料後，心想市面上現有飲料品牌眾多，如何才能脫穎而出，獲顧客青睞？該公司遂設計出一套實驗流程，然後找到 25 所中學進行實驗，每所中學的飲料配方僅有極微小的差異。各校均執行一檢定，取

H_0：喝此飲料無法提高記憶力，

H_a：喝此飲料能提高記憶力。

中學生升學壓力大，飲料若對提高記憶力有幫助，就有賣點。在 α=0.05 下，其中有一所學校得到顯著的結果。於是該公司以此校的結果(其餘 24 校的結果當然就不公開了)，完成一份研究報告，宣稱“經嚴格的統計檢定，證實喝該配方的飲料，能提高記憶力”，且大力促銷。一瓶才 20 元左右的飲料，會有那麼驚人的功能？不少人嗤之以鼻。但報告看起來，卻符合假設檢定的流程。

上述飲料提高記憶力的實驗，乃屬俗稱“德州神槍手謬

誤”(Texas sharpshooter fallacy)之一種，與中文裡的“先射箭再畫靶”類似。典故的由來，是美國有個德州人，朝著自己的穀倉(barn)射出多發子彈，隨即在彈孔最密集的區域畫一靶(shooting target)，然後自稱是神槍手(sharpshooter)。其後衍伸出，凡在大量的數據中，刻意挑選出對自己的觀點有利者，然後宣稱得到具統計顯著性的結果，至於其餘多筆對自己想要之觀點不利的數據，則棄之不用，皆可稱為“德州神槍手謬誤”。底下再來看一例子。

高壓電線可怕嗎？也就是對人體有害嗎？直至今日，仍屢有人對為輸送電力，而架於空中的高壓電線，深感疑慮。1992 年，瑞典曾有項研究，試圖找出空中上的高壓電線，對人們健康之影響。研究人員收集某地區高壓電線 300 公尺範圍內，所有居民的健康資料，時間長達 25 年。他們對超過 800 種大小疾病(ailments)，一一檢驗在該地區之發生率，是否較其他非住高壓電線地區，有顯著性的多。結果顯示，孩童白血病(leukemia)之染病率，為其他地區的 4 倍。依據此結果，研究主持者，施壓政府相關單位，須採取行動改善。只是當比較的疾病超過 800 種時，其中至少有一種疾病，其染病率較一般高達 4 倍，不過是隨機下正常會產生的現象。而後續的研究，果真不再發現高壓電線，與孩童白血病，有任何相關。

投擲一銅板 100 次，設得到 65 個正面，比公正銅板出現正面數之期望值 50，超出 3 個標準差，p-值=0.0026，所以對大部分常見的 α 值，此結果是顯著的，將拒絕銅板為公正之假

設。但若某中學的高二年級，有超過 4 百個學生，每位學生皆執行一投擲銅板 100 次之實驗，則就算每位學生拿的銅板皆為公正，其中若有學生擲出之正面數，與期望值 50，偏離至少 3 個標準差(≥65，或≤35)，絲毫不必訝異。

29 檢察官的謬誤

假設檢定雖是現代做決策之一重要依據，但我們已數度強調，見到一顯著事件發生，不宜立即見獵心喜，以為虛無假設可被推翻了。反而該抱著近鄉情怯的心理，仔細檢視這樣的結果，與前提(虛無假設)之成立或不成立，究竟有多大關連？若(自以為)依賴統計，過度重視顯著事件，有時會產生所謂“檢察官的謬誤”(prosecutor’s fallacy)。

來看個可能發生的情況。設有某法院審理某刑事案件，在無罪推定的前提下，虛無假設自然取為 H_0：被告無辜。假設開庭時，檢察官提出一疑點，有某現象 D 發生了，且指出若 H_0 為真，則現象 D 發生之機率僅為百萬分之 1，一個極小的 p-值。此時被告是否該俯首認罪？等等！現象 D 的發生，與 H_0 為真之關係大嗎？說不定不管 H_0 是否為真，D 發生之機率都是百萬分之 1。此時被告的律師，若有概念的話，不妨反問當 D 發生時，H_0 為真之機率又是多少？這其實是另一條件機率，須有更多的資訊才能求得。由於通常條件機率 $P(A|B)\neq P(B|A)$，故

$$P(D|H_0)=10^{-6}，$$

不表亦有

$$P(H_0|D)=10^{-6}。$$

因而檢察官企圖讓人相信的(底下 H_0^c 表 H_0 之餘集，在此即被告並非無辜)

$$P(H_0^c|D)=1-P(H_0|D)=1-10^{-6}\approx 1，$$

只能說是憑空冒出。例如，如前所述，若 D 與 H_0 獨立，且 $P(D)=10^{-6}$，則當然有 $P(D|H_0)=P(D)=10^{-6}$，但 $P(H_0|D)=P(H_0)$，可就不必然等於 10^{-6} 了。就算 D 之發生，的確屬於顯著事件，但若根本與 H_0 是否為真毫不相干，則提出 D 發生有何用？魚目混珠而已。只是在法庭上(當然政壇上也不少)，不少人屢會有意或無意地將 $P(D|H_0)$及 $P(H_0|D)$二者混淆，有時冤獄便因而產生了。由於這套利用條件機率的詭論，偶被檢察官引用來做佐證，故才會被稱為“檢察官的謬誤”。鑑於統計有時成了冤獄的幫凶，遂有人提出“靠統計數字定罪的危險”之警告，頗令統計人難堪。底下便以著名的“莎莉克拉克案”(The Sally Clark case)，來說明在法庭，若要引用統計來當證據，須格外謹慎。而另一方面，在辯論攻防時，若見到有人引用統計，也須提高警覺，不要迅即被“科學”說服。資料取自 Wikipedia(維基百科)。

英國的莎莉克拉克(Sally Clark，1964−2007，本名 Sally Lockyer，自結婚後以夫姓 Clark 取代娘家的姓 Lockyer)，是家中的獨生女，她父親是位資深警官，母親是位美容師。1990年，莎莉與同樣是初級律師(solicitor)的史提夫克拉克(Steve Clark)結婚。家庭及事業，一切看起來都很美好。1996 年 9 月，他們的老大誕生。這個健康的男嬰，不幸卻在當年 12

月，11 週大時，在家中猝死(Sudden cardiac death，縮寫 SCD，指突然的死亡)。莎莉好不容易才從悲傷中復原，於隔年(1997)11 月，又生了一個兒子，以為可開始過新生活了。豈料 8 週後，1998 年 1 月，嬰兒又在家中猝死，兩次兒子猝死，都是莎莉一人在家。只是人世的慘痛，豈僅是失去雙子？已難再從悲傷中復原的莎莉，面臨殺嬰的控訴。

起訴的檢察官，並沒有莎莉行凶的直接證據。但他就是認為，接連兩個嬰兒猝死，極不尋常，應為顯著事件。憑其直覺，檢察官完全不相信嬰兒猝死症(Sudden infant death syndrome，縮寫 SIDS，指嬰兒突然死亡，不論從其病史、身體檢查，或研究調查，都無法發現死因)，便是此二事件發生的真正原因。為了說服陪審團，這絕不是猝死，檢察官找來梅鐸(Sir Roy Meadow，1933−，他是位爵士)作證。梅鐸是位夙負盛名的小兒科醫生(paediatrician)，且上法庭作證的經驗豐富，他以"擅長"利用機率來佐證著稱。

梅鐸以簡易的方式，向陪審團說明，一家有 2 嬰兒接連猝死的機率有多小。他說同一家庭有兩個小孩死於 SIDS 的機率，為 7,300 萬分之 1。梅鐸承認這麼微小機率的事件，並不表示就不會發生。但他指出，這種意外，每 1 百年才會有 1 次。搞不清楚 7,300 萬分之 1 到底多小的人，於聽到百年 1 次，立即就懂了。人生不滿百，百年發生 1 次的事件，怎會在有生之年見到？因而此顯然不會是猝死。這是一記重磅！梅鐸還說，全英國兩個小孩的家庭，共有 1,500 萬個。對照 7,300 萬分之 1 的機率，這麼一聽又更明白了。不會發

生！又是一記重磅！既然不可能是猝死，“真相”便立即浮出了。陪審團接受梅鐸的證詞。1999 年，莎莉被判無期徒刑 (life imprisonment)，並於 2000 年入獄。直到 2003 年 1 月，經第二次上訴後，基於死嬰之新的病理報告出爐，最高法院改判莎莉無罪。只是遲來的正義，對莎莉已不具意義了。出獄後，莎莉一直處於精神不佳的狀態，有如槁木死灰，終日酗酒，茫然度日。2007 年 3 月，她因酒精中毒，死於家中。

我們先來看梅鐸的數據之由來。梅鐸宣稱，對一如克拉克這種家境良好且不抽煙的家庭，會發生 1 件嬰兒猝死(cot death)的機率為 1/8,543。因此這種家庭會發生 2 件嬰兒猝死的機率，為前述值之平方，即

$$(1/8{,}543)^2=1/72{,}982{,}849，$$

這是 7,300 萬分之 1 的機率產生之由來。那百年 1 次又是如何來的？梅鐸說,全英國每年約有 70 萬個新生兒,他將 7,300 除以 70，得 104.⋯，取近似值 100。應是 104 年發生 1 次，他還少算一些呢！只是將 7,300 除以 70，不知究竟代表什麼？

我們不能期望陪審團、檢察官或法官的數學、機率及統計，能有多好，那梅鐸的功力如何呢？身為醫生，能被封為爵士，總不至於浪得虛名吧！神秘的 1/8,543 這一嬰兒猝死機率,雖不知如何誕生,我們就先接受好了。但將兩個 1/8,543 相乘，就毫無根據了。同一家庭的 2 嬰兒先後猝死，真是獨立事件嗎？這對未曾謀面的兄弟 2 人，由於遺傳之關係，說

不定會有類似的基因缺陷。再加上照顧方式，及生長環境皆類似等因素，一家庭中的 2 猝死事件，絕不該在沒有依據下，就視為獨立。無論如何，不管三七二十一，就將兩機率值相乘，是很輕率的。不該是一個受過水準以上機率訓練的人，所會犯的錯。因此機率 7,300 萬分之 1，應是一極被低估下的值。至於因機率 7,300 萬分之 1，且每年約有 70 萬新生兒，將二者相除，就得到同一家庭兩件嬰兒猝死案，百年才會有一樁，這更是莫名其妙了。兩數字可以相除，但不表具有意義。只是既然有這麼多統計上的缺失，英國眾多統計學者，難道都不吭聲嗎？

首先，如我們之前所一再強調的，這裡犯了“檢察官的謬誤”。因就算 7,300 萬分之 1 的機率為正確，也不表在 2 嬰兒猝死下，莎莉無辜的機率也是 7,300 萬分之 1。這跟若某 A 君中樂透彩頭獎，不表他必是靠作弊得獎的情況類似。這是另一條件機率，要有更多的資訊，才能估算。另外，英國皇家統計學會(Royal Statistical Society，縮寫 RSS)，倒也沒有袖手旁觀，只是動作相當緩慢。學者總是這樣，不愛蹭熱度，在確定真有必要，且有極大的把握前，不會輕易出手。事實上，他們於 2001 年 10 月(莎莉入獄後隔年)，發表一公開的聲明，對本案裡的“法庭誤用統計”(misuse of statistics in the courts)，表示關切。並說“7,300 萬分之 1 的機率毫無統計依據”(“no statistical basis” for the “1 in 73 million”)。2002 年 1 月，RSS 還寫信給上議院大法官(Lord Chancellor)，明確指出 7,300 萬分之 1 的計算是錯的(the calculation leading to 1 in 73 million is invalid)。

再過 3 年，2005 年，英國醫學總會(General Medical Council，縮寫 GMC)，鑑於梅鐸曾多次在法庭上擔任專家證人，卻提供錯誤資訊，因而數度入人於罪，撤銷了他的醫師執照。雖經上訴後，隔年梅鐸重新拿回其執照，但名聲已毀了一大半了。最後，2004 年，索爾福德大學(University of Salford)的數學教授希爾(Ray Hill)，在期刊 Paediatric and Perinatal Epidemiology 上，發表一篇論文。他依據英國的統計資料，推導出 1 嬰兒猝死的機率約為 1/1,300，而非 1/8,543 那麼小。他並且估計出，一家庭若有 1 嬰兒猝死，則會再有 1 嬰兒猝死的機率，將提高 5 至 10 倍。看來梅鐸醫生自以為能善用統計，其實犯了不少錯，卻一直毫無所覺。統計！多少人假統計之名！

2016 年 7 月 5 日，中國時報有則標題是“16 年前性侵案 DNA 揪惡狼”之報導；相隔 3 天，同年 7 月 8 日，聯合報亦有則有關 DNA(deoxyribonucleic acid 之縮寫，去氧核醣核酸)比對的報導，標題是“警方如何鎖定爆炸案嫌犯？根據乘客傷勢”。今日警方屢能利用比對 DNA 來破案，福爾摩斯若復出，想必會嘆為觀止。

DNA 比對，成了今日破案的利器。而既然能讓 16 年前的犯案者，無所遁形，則不足為奇，也當然有平反冤獄的時候。有些人已含冤坐了多年牢，本再也不相信什麼老天有眼，結果拜 DNA 比對之賜，終於還了清白。由於有關 DNA

比對立功的報導，一再出現，使人們不禁以為，一旦祭出 DNA 比對，眾人便不得不臣服。大致是這樣沒錯，只是在運用時，如前兩則新聞內容所顯示的，仍需有其他佐證。即不能光是比對吻合，就立即以為案子破了，找到真相。若過度執著於 DNA 比對的高吻合率，有時便難免犯下“檢察官的謬誤”。底下給一實例。

2016 年 6 月號的 Scientific American (Volume 314，Issue 6)，有一篇標題為“When DNA Implicates the Innocent”的文章，作者是 Peter Andrey Smith。由林雅玲譯的“DNA 證據牽連無辜？”一文，則刊登在下一個月，台灣發行的“科學人”(2016 年 7 月號(No.173))。在該文中，先舉一 DNA 比對，引導錯誤的案例。2012 年 12 月，美國警方根據 DNA 的比對，指控一位名叫安德森(Lukis Anderson)的遊民，涉嫌矽谷富豪庫姆拉(Raveesh Kumra)的謀殺案。此控訴若被判有罪，最重是死刑。只是處在美國社會底層的安德森，與此案根本毫不相干，他有無懈可擊的不在場證明。在 11 月庫姆拉被謀殺的那晚之前，安德森便因醉酒且近乎昏迷(drunk and nearly comatose)，一直待在醫院接受治療。只是自以為依科學辦案的警方，見多了狡猾的嫌犯，當然不理會此對安德森有利的證據。那這件飛來橫禍，究竟是怎麼產生的？安德森的辯護律師團發現，安德森的 DNA，是在事發之後，隨著醫護人員進入庫姆拉家中。那幾位醫護人員，在當天稍早曾治療過安德森，然後在 3 個多小時後，不經意地把安德森的 DNA，帶到命案現場，並留在那裡，成為安德森涉案的“鐵證”。DNA

會移動？居然有這種事！有些人感到不可思議。其實倒也不必太驚訝，因我們已多次舉出不可能卻發生的事件了。

上述文章中指出：

> DNA 分析以統計模型做出預測，比起其他法醫技術，更為明確與客觀。(DNA analysis is more definitive and less subjective than other forensic techniques because it is predicated on statistical models.)

一般而言，利用統計，當然比其他方法更明確且更客觀，這是無庸置疑的。但該文要講的重點是：

> DNA 檢驗和其他證據一樣，透露的僅僅是完整案件的其中一個面向。…。如果我們過度依賴這個方法，以及在檢視 DNA 證據時，缺乏適度的質疑，誤判案例早晚會發生。例如，生物樣本可能分解，或遭到污染，而法官和陪審團可能曲解統計機率(statistical probability)。甚至，就像安德森一案所揭露的，皮膚細胞也可以“移動”。…。一個人攜帶的衣物，就算只是碰到另一人的脖子，都可能把後者的 DNA，轉移到此人從未接觸過的物體上。…。DNA 的轉移究竟有多常引發錯誤的指控，目前仍未知。

與 DNA 的比對類似，統計裡的假設檢定，此一科學的思維

與方法，的確常能讓人們將事情看得更清晰，因而有助於提高決策品質。但切記它並非萬無一失。若“過度依賴，而缺乏適度的質疑”，則誤判的發生，便絕非偶然。

30 統計顯著性

俗語說“狗咬人不是新聞，人咬狗才是新聞”。即事情若稀鬆平常，並不值得報導。至於不稀鬆平常，不就是顯著嗎？一向如此，不論統計裡的假設檢定，或媒體上的報導，具顯著性的事件，才會引起注意。我們說過，一事件是否為顯著，乃看其發生機率是否夠小。但我們也曾指出，小機率事件，只要觀測數夠多，或者說小機率碰到大樣本，其發生就不稀奇了。因而在 2300 萬人的台灣屬顯著事件，從全世界超過 79 億人(至 2021 年 12 月底)來看，可能便屬芝麻小事了；另外，一現今看來極為顯著的事件，若從幾千年人類歷史來看，也可能根本微不足道。

在台灣懷 5 胞胎便已極稀少了，若全都健康誕生，媒體必然大肆報導。2021 年 5 月 6 日有一則報導，非洲國家馬利(Republic of Mali)的一名婦女，於 5 月 5 日產下 9 胞胎，其中有 5 女 4 男，且全都平安。新聞裡說，婦女成功生下 7 胞胎已非常罕見，如今卻是破紀錄的 9 胞胎。之前的最多胞胎世界紀錄，是 2009 年 1 月 26 日，發生在美國 6 男 2 女的 8 胞胎。經過漫長 12 年多才新創的 9 胞胎紀錄，料想應可維持很久吧！沒想到才過 1 個月又 2 天，紀錄便又被打破了。南非(Republic of South Africa)有位婦女，於 2021 年 6 月 7 日，平安生下 7 男 3 女的 10 胞胎。不過其中有 5 個是自然生產，另 5 個則藉由剖腹。若加上家中原有的 6 個(包含一對雙胞

胎)，總計有 16 個孩子，這家將辛苦了。又，前述 8 胞胎及 9 胞胎，都是體外人工受孕(In vitro fertilization)，而這位生下 10 胞胎的媽媽，聲稱她是自然懷孕。不知如何求出，有人宣稱生下 6 胞胎的機率是 47 億分之 1，一個小的不得了之機率。會散發到世界各地媒體報導的事件，顯著性自然不必說，即其發生機率必定不是普通的小。但世界這麼大，各種顯著事件，仍層出不窮。

職業球賽裡，常有各種新紀錄誕生。以美國職業籃球 NBA 為例，共 30 支球隊，除特殊情況(如勞資糾紛或疫情等)外，每隊每年例行賽要打 82 場。已過世的傳奇人物布萊恩(Kobe Bryant，1978–2020)，曾立下不少輝煌的紀錄，像是 2006 年，他創下生涯單場最高的 81 分，是 NBA 史上單場得分第二高。即使這樣的頂尖球員，亦曾在 2012 年 3 月 31 日於主場，創下前 15 次出手皆落空的悽慘紀錄，那是他生涯最差的先發表現。底下我們來看，對布萊恩而言，連 15 投不進，究竟有多稀罕？

2013 年，布萊恩的阿基里斯腱斷裂(torn Achilles tendon)，雖健康復原，但在接下來的兩個賽季，又分別因膝蓋及肩膀傷患而報銷，2015–16 年的賽季他退休。後期受傷下的數據不計，依媒體報導，布萊恩在 NBA 的生涯，至連 15 投不進那場，共出賽 1,155 場，每場平均出手約 23.5 次，平均命中率則約為 42%。為了簡化，就假設布萊恩每場都出手 23 次，且每次出手的命中率為 42%。以 A 表布萊恩在一場比賽中，出現“至少連續 15 次出手落空”的事件，其機

率則以 $P(A)$表之。為何考慮“至少”？因並非只有恰好連 15 次出手全落空才會受到矚目，更多次連續出手落空，將會更令人訝異。A 包含連 16 次、連 17 次、…，及全部 23 次出手皆落空的事件。而一串連 15 次落空，可能發生在始自第 1 至第 15 次(出手)、始自第 2 至第 16 次、…，或始自第 9 至第 23 次，共 9 類情況。連 15 次落空後，不論進球或不進球，便都無妨，因都歸屬事件 A 了。但那連 15 次的前 1 次出手，必須是進球，否則連續落空的歸類，便要屬於開始更早。例如，對第 3 類情況，即始自第 3 至第 17 次連落空，其第 2 次出手一定是進球，不會是落空，但第 1 次出手則可以是進球也可以是落空。從第 2 類至第 9 類，那 8 類都是這種情況。至於第 1 類，則因始自第 1 次出手，故並無前 1 次。又因每次進球機率 0.42，落空機率 0.58，故得

$$P(A)=0.58^{15}+8\times0.42\times0.58^{15}\approx0.00123283908。$$

$P(A)$僅比 1000 分之 1 略大些，為布萊恩在特定的某場球賽裡，會發生連續 15 次出手皆落空的機率。但會為球迷及媒體所重視，當然並不限某場特定的比賽。我們已求出 $P(A)$，因此 A 未發生的機率為 1-$P(A)$，此值約為 0.99876716，很接近 1。由此得在布萊恩生涯 1,155 場的比賽中，事件 A 皆未發生之機率為

$$p=(1-P(A))^{1,155}\approx0.240554792。$$

故在 1,155 場裡，A 至少發生 1 次的機率為 $1-p\approx0.759$。超過 3/4 的機率，這一點都不算小了，可說就是等著看到布萊恩

連 15 投不進。最後，在相同的假設下，布萊恩發生“至少連續 14 次出手落空”的事件，其機率約為 0.932，比原先的 0.759 大了不少；若再放寬，改為連續 13 次出手落空，對布萊恩而言，這也算罕見了，此機率則約為 0.994，已相當接近 1 了。布萊恩外，NBA 當然仍有好幾位能與他相匹比的明星球員。若不限布萊恩，NBA 有某位頂級好手，曾至少連 15 次出手落空，幾乎可說是必然發生了。所以連 15 次出手皆落空，對某特定大高手，在特定的某場比賽，是相當罕見，但在其球員生涯，或擴大在整個 NBA 史上，有某位大高手發生，完全無須訝異。

大文豪莎士比亞(William Shakespeare，1564−1616)流傳下來的作品，包括 38 部戲劇、154 首十四行詩(Sonnet)，及一些長短詩歌，全部作品有 80 餘萬字。假設要一隻猴子坐在一打字機前敲打，一旦猴子不想打了，就換一隻打。若猴子打了 1 萬個字母，甚至 10 萬個，人們不會相信，其中會出現任何有意義的 5 個字以上的句子。但利用機率理論可求出，若持續不斷地打字，則“遲早”會出現一部連續且完整的莎士比亞全集。當然這“遲早”的時間，可能比我們所能想像的都長。雖發生時間有限(即發生的機率為 1)，但在我們有生之年，不必期望有可能看到。由於敲打出連續且完整的莎士比亞全集之機率為正(雖微乎其微)，由小機率碰到大樣本的概念，知只要時間夠長，莎士比亞全集便將出現。

包含學位論文及投稿論文，近年來不時爆出抄襲事件。雖有罪推定產生爭議，有些大學的研究所，仍訂出“研究生

學位論文專業符合及品質保證檢核作業要點”，其中規定研究生“應完成二階段論文原創性比對系統之檢核”，有些大學則要求研究生在畢業前，須先上傳論文，通過防抄襲的偵測軟體檢驗才行。抄襲比對系統，在接收到某論文後，會將該論文之內容，與資料庫內擁有的大量文件比對，然後迅速計算並標示出，可能抄襲的部分、可能的原始出處，及相似文字的百分比。有些防抄襲軟體，只要句子中有連續超過若干個字，與某文件中的內容相同，便會指出來。當然抄襲軟體列出後，仍須人為去檢視，才不會冤枉人。例如，論文中若引用某首詩，或某個定理，即可能有連續幾十個字，與某資料的內容吻合。通常只要在論文中，有講清楚引文的出處，便不至於被認為是抄襲。

幾年前，曾有部美國的電視影集，劇情主要是討論一律師事務所，為客戶打的各種官司。在某一集裡，有人委請律師，對抄襲他的書之出版社提出控告。上了法庭，被告律師答辯，依反抄襲軟體，兩書相似度僅 41%，以此佐證他的委託人並未抄襲，或者說即使有抄襲，也不算嚴重。劇中法官支持被告，同意 41%的相似比率並不高。原告及其律師當然都很不服氣。要知影印機若列印出的，與原件相似度為 95%，則影印機的品質不會被接受；影像與本人相似度，即使高達 99%，也會有人不滿意；但補習班考前猜題，命中率若能有 30%，就很驚人。所以相似度之高或低，應由發生的難易程度，即依發生機率的大小而定，不能由表面上的相似度百分比之大小來判定。但我們現在應知道了，兩書相似度若達

41%，必然是抄襲，絕不該被視為“如有雷同，純屬巧合”。

選舉若得票率差距過小，落選者常會不服，懷疑計票有問題。為讓落選者心服，有些選務機構遂規定，若因得票率(得票數除以有效票數)差距未超過千分之 3 而落選者，得聲請重新計票。換句話說，千分之 3 的差距被認為已夠小了，值得為此大費周章，將開票過程重走一遍。只是究竟怎樣的差距才算小？你知道當然不能用票數之絕對差來表示，因如果有效票數才 1 千，則相差 1 百票便很多了；而有效票數若達 1 百萬，則 1 百票的差異，便不算多。以得票率之差，來表示差異之大小，看起來較合理些，此為前述千分之 3 差距產生的背景。但以一固定的得票差異比率，來表示差距夠小之門檻，其實一點都不合理。我們已說過多次，以發生機率是否夠小，來反映差距是否夠小，才是合理的方式。

來看投擲一公正銅板 n 次的情況，且就假設 n 為偶數。你會預期必得到正反面數各半嗎？應該不會，因即使 $n=2$，得到 1 正 1 反的機率為 1/2，$n=4$ 時，有 1/4 的機率各得 2 正及 2 反。那 n 愈大，會愈容易得正反面數各半嗎？不會，反而更難！以 $n=10$ 為例，利用排列組合，可求出得到 5 正 5 反的機率為 252/1,024，小於 1/4。甚至可以證明，隨著 n 的增大，愈來愈不容易得正反面數各半，這點可能違反一般人的認知。底下以 S_n 表投擲 n 次投擲後，所得之正面數。對公正銅板，S_n 的期望值$=n/2$。所以平均來說，正反面數會出現各半。但對隨機現象，S_n 難免有偏差，通常不會剛好是 $n/2$。偏差有多大？S_n 的標準差為 $n^{1/2}/2$，此值隨著 n 之增大而增

大。即 n 愈大時，S_n 散佈的範圍將愈大，此範圍之寬度，以 $n^{1/2}/2$ 的“速度”成長。但相對偏差，即偏差比上投擲數 n，則愈來愈小。換句話說，所得正面數的比率 S_n/n，其期望值為 1/2，至於標準差 $1/(2n^{1/2})$，隨著 n 之增大而減小。即當 n 愈大，正面出現率 S_n/n 散佈在 1/2 附近，其範圍有愈來愈小的趨勢。因此 S_n/n 要偏離 1/2 一“固定距離”的機率將愈來愈小，這便是大數法則的內涵。大數法則針對的是正面出現率 S_n/n，而非正面出現的總數 S_n，這是有些人未能弄清楚的。

藉檢定銅板是否公正來說明。設投擲銅板 n 次，且仍以 S_n 表投擲 n 次後所得之正面數。先看 n=10,000。當銅板為公正，則 S_n 之期望值為 5,000，標準差為 50，正面數比反面數多千分之 3，即多 30，為 0.6 個標準差。此表得到正面數 5,015，反面數 4,985。在銅板為公正下，利用中央極限定理，出現正面數至少 5,015，比期望值多出 0.3 個標準差，其機率約為 0.3821，可說極易發生，因而不會拒絕銅板為公正。由此即知，設有 2 候選人，若有效票數為 1 萬，則兩人得票率差千分之 3，是一很小的差異，要求重計票算是合理。其次取 n=1,000,000。當銅板為公正，則 S_n 之期望值為 500,000，標準差為 500，千分之 3 的差距為 3,000，這相當於 6 個標準差，便不能等閒視之了。此表得到正面數 501,500，反面數 498,500。在銅板為公正下。利用中央極限定理，出現正面數至少 501,500，比期望值多出 3 個標準差，其機率約為 0.0013，可說極不容易發生，因此將會拒絕銅板為公正。由此即知，設有 2 候選人，若有效票數為 1 百萬，則兩人得票

率差千分之 3，此差異一點都不小，要求重新計票就沒有道理了。由以上討論知，得票率之差距究竟屬於大或小，乃與有效票數的多寡有關。有效票數若夠多，則千分之 3 的差距便已夠大，用統計術語來說，此差距屬於“顯著”，這時落敗的一方，即使心有不甘，也不該要求重新計票。附帶一提，不難看出，並非將重新計票條件中的得票率差距縮小，如改為未超過千分之 2，問題便解決。而讀者現在想必亦可了解，做統計分析時，何以會在意樣本大小，因取樣愈多偏差將愈小。

2004 年中華民國總統選舉主要有 A、B 兩組候選人。開票後，A 組得 6,471,970 票，B 組得 6,442,452 票，總有效票數為 12,914,422，兩組得票率分別為 50.114283%，及 49.8857169%。得票數相差 29,518 票，得票率相差約 0.228566%。差不到 3 萬票，且得票率差不到 0.23%(小於千分之 3)，落敗的 B 組，百般不服，應不待開召開“專家會議”，便已想要求重新計票了。在檢定銅板是否公正，且 n=12,914,422，於銅板為公正下，S_n 之標準差約為 1,796.83。當正反面數差了 29,518 個，即出現的正面數，較公正銅板預期出現的次數，多出 14,759 個，即多出約 8.2139 個標準差。此機率就不給了，只要想一般偏離 3 個標準，就已不得了了，如今偏離超過 8 個標準差，豈能視為偶然？當然不會接受銅板為公正。故雖得票率差不到 0.23%，小於千分之 3，但此時雙方的差異，是相當顯著的，根本不該要求重新計票。

31 估計

什麼是“信史”？有些人以為乃指“較為可信的歷史”。只要較為可信？看起來似乎不太嚴格。這是因即使發生在今日的事，常也撲朔迷離，眾說紛紜，難以確定真相，何況是幾百幾千年前的事？因而只要“較為可信”，便接受是信史了。條件這麼寬鬆，那信史不就很多嗎？是沒錯，不過史學界的態度普遍更嚴謹。他們對信史的要求是，“有同時期的文字記載，且有文物遺跡佐證的歷史。”這定義雖嚴格不少，但其實相當合理，也是我們所接受之定義。

中國的信史始於何時？夏朝嗎？不！商朝！即使如此，商朝的年代至今並無定論。有些文獻只概略地說從西元前 17 至 11 世紀；有些則明確地指出，從西元前 1562 年，至西元前 1066 年，長達將近 5 百年。但夏朝不見了！昔日中小學時，朗朗上口的中國朝代，“黃帝唐虞夏商周，秦漢三國晉，…”，以及講到中國的道統時，“堯舜禹湯，文武周公孔子，…”，結果不要說黃帝及堯舜都可能是虛擬，還連從大禹開始，一整個夏朝(傳說中約自西元前 2070−1600 年)，都被視為子虛烏有？這樣中國的歷史，不就只剩約 3 千 5 百年？比起距今約 6 千年前的兩河流域文明，及至少有 5 千年歷史的埃及文明，中國成了小老弟了。夏是中國史書所載，第一個世襲制的朝代，意義重大。“左傳”裡有“以服事諸夏，楚食華夏，商不謀夏，夷不亂華。”顯示至少自西周起，

便有諸夏、華夏、夏及華的稱呼。我們以華夏民族自居，以華夏文明自豪，華夏與中華差不多同義，夏常被視為中國的代名詞。大禹治水、少康中興，及商湯滅夏等，都是大家自幼起，耳熟能詳的故事。佔如此獨特地位的夏朝，怎麼卻如女媧補天，及黃帝大戰蚩尤一般，不過是故事，完全不能當真？只是在“論語”跟“孟子”裡，孔子(西元前 551–479 年)及孟子(西元前 372–289 年)，都曾多次提到“禹”，司馬遷(西元前 145 年–？)的“史記”中，亦有“夏本紀”，後世史書也屢屢有關夏朝的記載，這些難道都不算嗎？

自周朝以來，文獻中關於夏的記載不少，但至今尚未發現在“傳說中”的夏朝時期，任何有關夏朝，或就僅是禹的文字記載。從 19 世紀末起，西方考古學(Archaeology)的相關概念傳入中國。過去 1 百多年來，整個大陸地區，雖不時有古物出土，但卻從沒有夏朝的，連 1 件也沒有。即使在“傳說中”夏朝之後的商朝，雖自 19 世紀末以來，有大量震驚世人的甲骨文出土，讓今人對商朝更加了解，惟其中卻缺乏任何與夏朝或大禹有關的記載。有人可能認為，就算沒有考古的支持，孔子、孟子及司馬遷講的，總可以相信吧！不行！要知人們連對父母說他們以前如何如何，有時都難免半信半疑，何況孔子及孟子，距“傳說中”的禹，都有 1 千 5 百年以上，司馬遷就更晚了。因而並無法以孔孟不時提到，且百般推崇大禹，以及我國第一部“正史”中之記載，來佐證夏朝及禹的存在。原來我們自豪的中華 5 千年文化，得打個 7 折。消失的夏朝，少掉的 4、5 百年，令人沮喪！

2016 年 8 月 5 日，國立台灣大學，發佈一則乍看之下，令人振奮的新聞，標題是“人類學系高德教授證實大禹治水及建立夏王朝榮登 Science(科學雜誌)”。原來該校人類學系的高德(David J. Cohen)教授，其所參與的跨國研究團隊，在著名的 Science 雜誌，於 2016 年 8 月 5 日那期，刊登 1 篇在大陸青海省的地質調查報告，提供黃河流域在西元前 1920 年，爆發大洪水的證據。在該篇論文中估計，巨大的水流量，其影響遍及黃河下游及其支流約 2 千公里，並可能造成黃河原先的天然防洪堤潰堤，使得河水大量溢出黃河河道。比對大洪水發生的時間點、威力及所造成的影響，該論文推斷，此事件提供中國古代文明起源，“傳說中”有關“大禹治水”，及建立夏朝的故事之基礎。論文中也探討，考古學上所見大洪水與社會政治轉變的相互關係，比對與史學上的紀錄是否能相呼應。論文中且指出，若夏朝的建立，與歷史上記載的大洪水確實相關，且本篇論文所討論的大洪水事件，在程度與時間上，皆符合“書經”及“史記”裡的相關記載，則便能確認西元前 1 千 9 百年為禹建立夏朝的年代。其中提到的“書經”即“尚書”，主要記載發生在夏、商、周的一些重要事蹟。

上述研究之推論雖引人注意，只是如何確定夏朝的建立，與歷史上“記載”的大洪水有關？論文中似便沒說了，畢竟一切都是傳說。像這樣關於約 4 千年前，且涉及 2 千公里外的估計，就算發表在頂尖雜誌，恐仍無法取信史學界大禹確實存在。

處在此一隨機世界，自古以來，人們可說經常在估計。估計的範圍相當廣泛，只要自己未知便可估計。估計可以憑空或隨意地猜，如拿起一個西瓜，敲一敲，估計它甜不甜；對初認識的朋友，上下打量，估計他的年齡。估計也可有套程序，如前述估計大禹的存在。另外，亦可有不算淺顯的計算。如早在西元前 235 年，旅居埃及的一位希臘地理學家埃拉托斯特尼(Eratosthenes，西元前 276−194 年)，便以幾何學的推導，求出地球的周長約介於 39,690 公里至 46,620 公里間。此估計並非太離譜，因經過兩極的地球實際周長約為 40,008 公里。其他如月球及太陽的大小、地球與月球間的距離等，當時也都能估計出。估計也可很超乎現實，如時至今日，仍持續有科學家在估計地球的年齡，雖不同人的估計，其差異可長達數十億年，且極可能永遠不會知道正確值，但科學家依然樂此不疲，每隔一陣子，便有一新的估計值產生。

對不同的情況，人們常需要做估計，只是該如何估計呢？曾有新聞報導，日本九州佐賀縣有一神社，依據稀飯發霉的情況，及霉菌的顏色，以估計運勢。不僅如此，還能估計何時會有地震發生。大陸揚州有位算命師，預測“耳中長毛”者，可活過 99 歲(見 2006 年 2 月號“科學人”雜誌“總編輯的話”)。究竟該如何估計？對一般人，可以很有道理也可毫無道理。但對統計學家，估計應有些依據。給出估計後，便屢會用以對未來做預測。統計學家的預測，總該比用亂猜，或諸如依稀飯、水晶球來判斷等，更準確才對。而合理的估計法，通常也會具備一些優點。但怎樣才算是好的估

計？有兩個鐘，一個停止不動，永遠顯示同一時間，另一個則每日慢 1 分鐘，何者較準？有人以為是前者，因它每天有 2 次顯示正確時間，後者則每 720 天(將近 2 年，因要慢 12 小時，即 720 分鐘)，才有一次顯示正確時間。故時鐘準確的定義為何，得先給出來。正如選美比賽前，也得先規定評比標準，否則無從排序。所以估計的評比，也得先給出標準。

人們常說眼見為實，但真的是這樣嗎？不知大家是否有下述這類經驗：去餐廳吃飯，經常是座無虛席，看來景氣極佳。但餐飲業者，卻常抱怨生意不好做。這究竟是怎麼回事？假設有家餐廳，有 100 個座位。週一至週五，每晚只有 10 個人去用餐，而週六及週日，兩晚皆客滿，若未預訂是沒有座位的。如此一週 7 個晚上，共有 250 個人去用餐，平均每晚只有 $250/7\approx35.71$(人)，即不到 36 個客人，生意的確不能說很好。但有 80%(=200/250)的顧客，其印象是餐廳生意真好。再看一例。某大學大一學生，有不少抱怨通識課程難選，總是班班爆滿。但實際上卻有很多班通識課程，修課人數並不太多，這又是怎麼回事？假設共開設 20 班通識課程，每班修課人數之上限皆為 100。選課結束後，4 班有 100 個人選，另 16 班各只有 10 人選。20 班選課總人數共 $4\times100+16\times10=560$(人)，平均每班只有 $560/20=28$(人)，並不算多。但說不定那 $400/560\approx71.43\%$的學生，是早早等待，系統一開放選課，便立即上網選，雖搶到仍心有餘悸。而若對這 20 班共 560 個學生發問卷，調查其修課的班上有多少人？又假設 560 張問卷皆收回，則其中將有 400 張填 100 人，160 張填 10 人，

總共所得修課人數有 400×100+160×10=41,600(人)。問卷顯示，平均每班有 41,600/560≈74.29(人)，與前述每班實際平均只有 28 人差異不小。這類例子很多。如旅遊名勝景點，遊客常抱怨這麼賺錢，卻不提高服務品質。如停車位太少不說，連餐廳都不多設幾處。真實的情況是，除了少數假日遊客摩肩接踵外，其餘的日子，大多是門可羅雀。若遊客由其親身體驗，以估計該景點平均每天有多少遊客，則大部分人的估計，可能極不正確。

有時估計錯誤，或不重視估計，將可能造成極大的損失。民國 103 年 11 日 14 日，聯合晚報有一則標題是“教部次長陳德華：當初沒料到少子化”之報導：

> 教改滿 20 年，當年廣設高中大學的決策，造成現在大學面臨轉型退場苦果。教育部次長陳德華表示，當初教改提出廣設高中大學訴求時，並沒有少子化問題；民國 87 年人口出生數突然下降後，教育部開始限縮大學增加分部及系所，但“煞車已來不及”，教育部預計年底提出高教創新轉型行動方案。…。陳德華坦承當年行政院教改諮議報告書，並未列入少子女化可能造成的衝擊；但他也說，在民國 83 年以前，台灣一年出生人口數有 32 萬至 34 萬人，沒有少子女化現象。少子女化現象直到民國 87 年才出現，…。

沒料到少子化？事實上，早先台灣每年的新生兒數，都在 40

萬上下，像是 1978 至 1982 年，年年超過 40 萬。1982 年有 40.5 萬，但那是台灣年新生兒數，最後一次達到 40 萬，之後便開始下降了。1994 年 4 月 10 日 410 教改大遊行的前 5 年，即 1989 至 1993 年，每年新生兒數分別約為 31.5 萬、33.5 萬、32.1 萬、32.1 萬，及 32.5 萬，皆較 1982 年，少掉 7 萬以上。因此即使什麼也不做，依此趨勢，都將自動小班小校了，且升學壓力亦將逐漸緩和。怎須上街頭，爭取落實小班小校，及廣設高中大學？每年 32 萬左右的新生兒數，維持到 1997 年，隔年(虎年)掉到約 27.1 萬，較前一年驟降約 5.5 萬。因新生兒數下降，加上大學錄取名額增多，自然使重考生減少，因而大學錄取率必然逐年提高。這是極簡單的道理，不須用到什麼深奧理論。如果連這種情況都無法掌握，更複雜的情況豈有辦法不離譜的估計？附帶一提，2010 年(虎年)的新生兒數，少到約 16.6 萬，那時已覺很驚人了；2020 年的新生兒數，為 16 萬 5249，而 2021 年的新生兒數，估計將不足 15 萬，創下歷年最低紀錄。大學逐漸退場，已是進行式了。這是不重視估計的後果之一例。

32 估計方法

我們先看點估計(point estimation)，即以一個統計量，來估計一參數。

下課時學生嘻鬧間，教室的玻璃窗被打破了。誰是禍首？無人承認，這時老師往往從平日最調皮的同學開始詢問。社區有竊案發生，警方常從有前科且具地緣關係者開始追查。怎可懷有偏見？沒辦法，若優先約談向來規矩的學生，或毫無前科者，豈不更不合理？事實上，醫生看診，常也是從病人的症狀，推測那一種病最易產生此症狀。會誤判嗎？當然會。這類例子生活裡處處可見。看電影時，人們屢推測下一步會如何，才會有之前的劇情。編劇會特意誤導觀眾嗎？不無可能，因畢竟有時並不想讓觀眾輕易猜中下一步，有時會製造些懸疑。但劇情總是較常有個脈絡可尋，很少會從頭到尾一直跳躍不已。

像這種由觀測到的結果，推測最可能使此結果發生的原因，是人們常有的思維。由此便衍生一種估計法。設有一隨機現象，知其發生結果之機率密函數，但其中含有某未知參數，擬估計該參數。如獨立地投擲一出現正面機率為 p 之銅板 n 次，其中 $0\leq p\leq 1$，以 X_1，X_2，…，X_n 分別表各次所得之結果，其中 $X_i=1$，或 0，就依第 i 次得到正面或反面，$i=1$，…，n。令 $S_n=X_1+X_2+\cdots+X_n$，且 $s_n=x_1+x_2+\cdots+x_n$。因 X_i 有 $Ber(p)$ 分佈，故 X_1，X_2，…，X_n 之聯合機率密度函數為

$$f(x_1,x_2,\cdots,x_n|p)=p^{s_n}(1-p)^{n-s_n}。$$

今欲估計參數 p，從所得觀測值 x_1，x_2，…，x_n，推測究竟 p 為何，才會使 x_1，x_2，…，x_n 被觀測到之機率最大？這是一種常用的估計法，在統計學裡稱為最大概似法(method of maximum likelihood)，所得之估計量值，稱為最大概似估計值(maximum likelihood estimate，縮寫為 MLE)；所得之估計量，稱為最大概似估計量(maximum likelihood estimator，縮寫亦為 MLE)。

現在來看對前述問題，最大概似估計值如何求出？將 $f(x_1,x_2,\cdots,x_n|p)$視為一 p 之函數，而將 x_1，x_2，…，x_n 固定(因 x_1，x_2，…，x_n 已被視為觀測值，固定下來，如今會變的是 p)，且以 $L(p|x_1,x_2,\cdots,x_n)$表此函數，函數之形式與 $f(x_1,x_2,\cdots,x_n|p)$相同，只是變數不同，換成 p。即

$$L(p|x_1,x_2,\cdots,x_n)=p^{s_n}(1-p)^{n-s_n}，0\le p\le 1，$$

並稱此為概似函數(likelihood function)，也就是可能性函數。我們想求 p，使得 $L(p|x_1,x_2,\cdots,x_n)$最大。也就是想求 p 的函數 $L(p|x_1,x_2,\cdots,x_n)$，其極大值發生在何處？經由微分，得極大值發生在 $p=s_n/n$。即 p 之最大概似估計值為 s_n/n，且 p 之最大概似估計量為 S_n/n。我們便以統計量 S_n/n 來估計 p。

再看另一估計法。對可重覆觀測的實驗，就仍以前述投擲銅板的情況為例，由於當 X 有 $Ber(p)$分佈時，X 之期望值 $E(X)=p$，我們遂以樣本平均 S_n/n 取代期望值 $E(X)$，即用樣本

平均 S_n/n 來估計 p。為什麼會想到以樣本平均來取代期望值 $E(X)$？因由大數法則，當 n 很大時，S_n/n 大致會在 p 附近一小範圍內波動。即在 n 很大時，S_n/n 於某種意義下，會很接近 p。也就是 n 很大時，以 S_n/n 來估計 p，不致於會有太大的偏差。由於在統計裡，$E(X)$稱為 X 之一次動差，$E(X^2)$稱為 X 之二次動差，餘類推，故這種估計法，稱為動差法(method of moments)，而得到之估計量，便稱為動差估計量(method of moments estimator)。有大數法則支持，有些學者對於可重覆觀測的現象，凡估計或推論，傾向依賴相對頻率，這樣的思維，被歸類為頻率學派(frequentist)。19 世紀時，由於崇尚大數法則，頻率學派可說主導統計的思維。依如此思維的推導，稱之為古典的作法(classical approach)。採用此法，通常只要樣本愈多，對參數的推估就愈精準。

對前述投擲銅板之例，以不同的兩個方法，動差法及最大概似法，得到相同的估計量，這並不奇怪，因畢竟這是兩個很合理的估計法，常會英雄所見略同。但當然兩種方法，也會有得到不同估計量的時候。底下給一例。

設 X_1，X_2，…，X_n 為一組由機率密度函數 $f(x|\theta)=1/\theta$，$x=1$，…，θ，所產生之隨機樣本，其中 θ 為一未知正整數。可看出這組隨機變數，其共同分佈是離散型的均勻分佈 $U[1,\theta]$。一個例子是，袋中有若干張紙牌，每張上有一數字，各張上的數字皆不相同，只知分別是 1，2，…，但不知最大為何。依序取牌，每次取出後放回，n 次後得到數字 X_1，X_2，…，X_n。現想據此估計數字的上限 θ。顯然 $E(X)=(\theta+1)/2$。

以 S_n/n 取代 $E(X)$，得 θ 之動差估計量為 $2S_n/n-1$。由於 X_1，X_2，⋯，X_n 都不可超過 θ，而樣本平均 S_n/n 可以很小，當 X_1，X_2，⋯，X_n 中，有比 $2S_n/n-1$ 大者，此估計量便不合理了。例如，設 $n=10$，且觀測到的 X_1，X_2，⋯，X_{10} 中，有 9 個 1，及 1 個 10，則 $S_n/n=1.9$，且 θ 之動差估計值為 $2\times1.9-1=2.8$。θ 明明須為正整數，如今以一非整數 2.8 去估計它，就已令人不安了；而 X_1，X_2，⋯，X_n，明明都不可超過 θ，如今卻在得到一觀測值 10 之下，還以 2.8 去估計 θ，更令人覺得這估計值實在不妥。但這卻被認為是由一不錯的動差法所得之估計值。此處樣本數 $n=10$ 並不大，當樣本數太小，估計值有時會不盡如人意。其次來求最大概似估計量。X_1，X_2，⋯，X_n 之聯合機率密度函數為

$$f(x_1,x_2,\cdots,x_n|\theta)=1/\theta^n，1\leq x_1,x_2,\cdots,x_n\leq\theta。$$

故概似函數為

$$L(\theta|x_1,x_2,\cdots,x_n)=1/\theta^n，1\leq x_1,x_2,\cdots,x_n\leq\theta。$$

上述概似函數之極大值發生在何處？顯然 θ 要愈小愈好。但又須滿足 $\theta\geq x_i$，$i=1$，⋯，n。故取 θ 為 x_i，$i=1$，⋯，n，中之極大值即可。現令 $X_{(n)}=\text{Max}\{X_1，X_2，\cdots，X_n\}$ 表 X_1，X_2，⋯，X_n 中之最大的量，且以 $X_{(n)}$ 估計 θ。此 θ 之最大概似估計量為 $X_{(n)}$ 必為正整數不說，且較動差估計量 $2S_n/n-1$ 合理多了。當然若 n 很大時，X_1，X_2，⋯，X_n 中，由大數法則，S_n/n 將有高機率會很接近 $E(X)=(\theta+1)/2$。於是 $2S_n/n-1$ 將有高機率會很接近 θ。如此一來，X_1，X_2，⋯，X_n 中，會有比 θ 大者，其

機率就很小了。不過不論 n 多大，$2S_n/n$-1 都不一定是整數。

估計法並不限上述兩種。門診時，聽完病人的描述後，醫生開單要病人去做某些項檢查。醫生對病因已有判斷，不外 A、B 及 C 等 3 種，機率各為 0.4、0.3 及 0.3。檢查完，看到報告後，醫生的判斷有了改變，病因 A 及 B 兩種，機率各改為 0.8 及 0.2，C 則不可能。對於分佈中的參數 θ，常被視為一未知但固定的量，由所獲得之隨機樣本 X_1，X_2，…，X_n，來估計 θ。但有時對參數會有些事前的看法，也就是不視為未知且固定，而當做一隨機變數。例如，當估計銅板出現正面的機率 p 時，以為既然是政府發行的，即使不是公正(p=0.5)，也不致於太偏頗，遂取 p 有 U[0.45,0.55]，即 p 在區間[0.45,0.55]均勻分佈。此處 U[0.45,0.55]，便稱為 p 之事前分佈(prior distribution)。由於事前分佈可以很主觀，因而常就被視為估計者之主觀的分佈(subjective distribution)。在此所謂主觀，並非一負面的描述，因此分佈的產生，常也是基於過去的資料，並非就是隨意猜測。由於主觀分佈是在實驗觀測前就建立的，故才稱為事前分佈。取樣後，修正對參數 θ 分佈之看法，便得 θ 之事後分佈(posterior distribution)，並據此估計 θ。這種方法，稱為貝氏作法(Bayesian approach)，持此主張者，便稱貝氏學派(Bayesian)。貝氏一詞之由來，是因在修正分佈時，用到條件機率裡的貝氏定理。在貝氏作法裡，一個常用來估計 θ 的量，為事後分佈的期望值，即 $E(\theta|X)$，且稱之為貝氏估計量(Bayesian estimator)。底下給一簡單的例子。

假設以抽血檢驗 D 君是否患有某特定疾病。令 X 表檢驗的結果，X=1 或 0，分別表 D 君檢驗後呈正反應及負反應；又令 θ 表 D 君有病或無病的狀態，θ=1 或 0，分別表 D 君有病或無病。又設 X 之機率密度函數為

$$f(1|\theta=1)=0.8，f(0|\theta=1)=0.2，f(1|\theta=0)=0.3，f(0|\theta=0)=0.7；$$

現將 θ 之事前分佈設為

$$\pi(1)=0.05，\pi(0)=0.95。$$

檢驗前 D 君有病的機率為 0.05，我們想利用貝氏作法，以估計檢驗後 D 君有病的機率。由假設，得 X 之邊際機率密度函數為

$$m(1)=P(X=1)=f(1|\theta=1)\pi(1)+f(1|\theta=0)\pi(0)=0.04+0.285=0.325，$$

$$m(0)=P(X=0)=f(0|\theta=1)\pi(1)+f(0|\theta=0)\pi(0)=0.01+0.665=0.675。$$

故在給定 X 之下，θ 之事後分佈為

$$\pi(1|1)=f(1|\theta=1)\pi(1)/m(1)=0.04/0.325\approx 0.123，$$

$$\pi(0|1)=f(1|\theta=0)\pi(0)/m(1)=0.285/0.325\approx 0.877。$$

$$\pi(1|0)=f(0|\theta=1)\pi(1)/m(0)=0.01/0.675\approx 0.0148，$$

$$\pi(0|0)=f(0|0)\pi(0)/m(0)=0.665/0.675\approx 0.9852。$$

故 θ 之貝氏估計量 $E(\theta|\mathrm{X})$，對 X=1，及 0 時，分別為

$$E(\theta|1)=P(\theta=1|1)=\pi(1|1)\approx 0.123，$$

及

$$E(\theta|0)=P(\theta=1|0)=\pi(1|0)\approx 0.0148。$$

即 X=1(檢驗呈正反應)時，以 0.123 估計 $P(\theta=1)$，X=0(檢驗呈負反應)時，以 0.0148 估計 $P(\theta=1)$。我們發現，當檢驗呈正反應時，D 君會有病的機率之貝氏估計值約為 0.123，並不大；而當檢驗呈負反應時，D 君會有病的機率之貝氏估計值約為 0.0148，就更小了。其中關鍵是因在人口中，會有病的機率只有 0.05，本來就不大。

對上例，假設有一組隨機樣本 X_1，X_2，…，X_n，即 D 君重複檢驗 n 次,則最大概似估計量及動差估計量,分別為何？在此二情況下，當然都不考慮 θ 之事前分佈了。由於 X_1，X_2，…，X_n 之聯合機率密度函數為

$$f(x_1,x_2,\cdots,x_n|\theta=1)=0.8^{s_n}0.2^{n-s_n}，$$

$$f(x_1,x_2,\cdots,x_n|\theta=0)=0.3^{s_n}0.7^{n-s_n}，$$

其中如前 $s_n=x_1+x_2+\cdots+x_n$。則概似函數為

$$L(\theta|x_1,x_2,\cdots,x_n)=0.8^{s_n}0.2^{n-s_n}，\theta=1，$$

$$L(\theta|x_1,x_2,\cdots,x_n)=0.3^{s_n}0.7^{n-s_n}，\theta=0。$$

故最大概似估計量為

$(28/3)^{S_n}(2/7)^n>1$ 時，估計 $\theta=1$，

$(28/3)^{S_n}(2/7)^n\leq 1$ 時，估計 $\theta=0$。

至於動差估計量，自然便是以 S_n/n 估計 $P(\theta=1)$了。

點估計的方法並不僅上述幾種，此處僅是初步的介紹，並未涉及太複雜的估計量。

33 估計量之評比

不同的估計法，有時會導致不同的估計量，如何評比各估計量之優劣？在此來看點估計。首先，若不設準則，便無所謂好的估計量，遑論最佳估計量？舉個例子來看，假設欲估計袋中紅球所佔比例 p。有人就隨機取樣 n 次，每次取出後放回，得到 X_1，X_2，…，X_n，分別表各次所得之結果，其中 $X_i=1$，或 0，就依第 i 次得到紅球或非紅球，$i=1$，…，n。令 $S_n=X_1+X_2+\cdots+X_n$，$n\geq 1$。眾所皆知，樣本平均 S_n/n 為一常見之 p 的估計量。但不論 n 多大，S_n/n 都不一定會很接近 p，當然 n 很大時，不接近的機率很小。不過，若某君不取樣，就隨興地以 0.48 來估計 p，亂猜的估計量應是很荒謬吧？或保守一點地說，該比 S_n/n 差吧！倒也未必，S_n/n 不一定等於 0.48，但若 p 真的是 0.48，該君便估計正確了。

某家有 3 個小孩，在分配物品時，很難每次每個小孩拿到的分量皆相同。對於少給的，媽媽承諾下次會給他多些。此家孩子講理，幾次下來每人平均拿到的若一樣多，便能接受，不會計較了。另外，某委員會中有位委員，開會遲到時，屢以時間沒估準為藉口，幾次下來，其他委員提醒他，若時間沒估準，應有時會早到，不該次次晚到。上二例顯示，人們常有平均須是準的之概念，底下我們要介紹的不偏估計量 (unbiased estimator)，可說便是這樣產生的。設有一組 n 個隨機樣本 X_1，X_2，…，X_n，並以統計量 $T=T_n(X_1$，X_2，…，$X_n)$

來估計某未知參數 θ。若滿足

$$E(T)=\theta，$$

則 T 便稱為 θ 之一不偏估計量，或就簡單地說 T 是不偏的(unbiased)。如前，此表有時高估，有時低估，但平均會是準的。直觀上，不偏性(unbiasedness)似乎是“好的”估計量該具備的條件，但其實有些不偏估計量，卻存在明顯瑕疵。又若 T 非不偏估計量，便稱 T 為偏差估計量(biased estimator)，或就說 T 為偏差的。

一般而言，以一估計量 T 來估計參數 θ，會希望 T 與 θ 之差距，即$|T-\theta|$要愈小愈好。但 $T-\theta$ 與未知的參數 θ 有關，大小無從知道；而且 $T-\theta$ 為一隨機變數，其大小依賴觀測的隨機量 T 而定，所以$|T-\theta|$無從曉得到底多大。又在數學裡，有關絕對值的處理較麻煩，不妨想想$|f(x)|$在區間$[1,10]$之積分，其中 $f(x)=x\sin x-\log x$，要決定 $f(x)$何時為正，何時為負，以去掉絕對值符號，實非易事。退而求其次，我們考慮均方差(mean squared error，簡稱 MSE)$R(\theta,T)$，其定義為

$$R(\theta,T)=E((T-\theta)^2)。$$

將上式改寫如下

$$R(\theta,T)=E((T-E(T)+E(T)-\theta)^2)$$

$$=E((T-E(T))^2)+E((E(T)-\theta)^2)+2E((T-E(T))(E(T)-\theta))$$

$$=\mathrm{Var}(T)+b^2(\theta,T)，$$

我們便有

(1) $R(\theta,T)=\mathrm{Var}(T)+b^2(\theta,T)$，

其中

(2) $b(\theta,T)=E(T)-\theta$，

稱為 T 之偏差(bias)。此處用到乘積公式

$$(a+b)^2=a^2+b^2+2ab，$$

且因 $E(T)-\theta$ 已非隨機變數，故

$$E((T-E(T))(E(T)-\theta))=(E(T)-\theta)E((T-E(T))=0，$$

其中用到 $E((T-E(T))=E(T)-E(E(T))=E(T)-E(T)=0$。

我們知道，期望值有如一隨機變數分佈之一中心。上述偏差 $b(\theta,T)$，便是量測估計量 T 之中心 $E(T)$，與欲估計的參數 θ，兩者間偏差之大小。因此 MSE 可分成兩部分 $\mathrm{Var}(T)$ 及 $b^2(\theta,T)$，前者即估計量 T 之變異，可顯示精準性；後者則用來描述偏差之大小，以顯示正確性(accuracy)。以射箭為例。若射在靶上的點都很接近(表 $\mathrm{Var}(T)$較小)，可說射得很穩定，相當精準。而若這些點之集中處偏離靶心(即 $b(\theta,T)$不小)，便是正確性不夠。至於若箭在靶上各處(或靶外)散佈，便是既不精準又不正確(表 $R(\theta,T)$較大)。理想狀態當然是不但$|b(\theta,T)|$愈小愈好(0 是最小的)，且 $\mathrm{Var}(T)$也愈小愈好。底下給個簡單的例子。

欲估計一銅板出現正面的機率 θ，重複投擲後，得到 X_1，X_2，…，X_n，分別表各次所得之結果，其中 $X_i=1$，或 0，就依第 i 次得到正或反面，$i=1$，…，n。人們常取 $T_n=S_n/n$ 做為 θ 之估計量，因 $E(T_n)=\theta$。故

$$b(\theta,T_n)=E(T_n)-\theta=0，$$

且

$$\mathrm{Var}(T_n)=\mathrm{Var}(X_1+X_2+\cdots+X_n/n)=n\mathrm{Var}(X_1)/n^2=\theta(1-\theta)/n。$$

即得

$$R(\theta,T_n)=\theta(1-\theta)/n。$$

可看出對每一 $n\geq1$，T_n 為 θ 之一不偏估計量、隨著樣本數 n 的增大，MSE 愈來愈小，且 $n\to\infty$ 時，$R(\theta,T_n)\to0$。

設參數 θ 有二估計量 U 與 V。若對每一可能的 θ，皆有 $R(\theta,U)\leq R(\theta,V)$，且至少有一 θ，使得嚴格不等式成立，此時我們說 U 較 V 為佳(U is better than V)，且稱 V 為不可採用的(inadmissible)。一估計量 U，若不存在較其為佳之估計量，便稱為可採用的(admissible)。此處所謂可採用與不可採用，乃以 MSE 為評比標準。做決策時，“可採用的”之概念用途廣泛。例如，女孩在找對象時(就依自己所訂的標準)，在周遭可挑的人選中，想找到最佳固然不易，因人總是各有優缺點，但無論如何，要選取可採用的。少有女孩會看上樣樣不如人者。像是女孩在意的因素設有 100 個，若 A 君每項因

素皆不如 B 君，則 A 君當然該被淘汰了。處在一團體(或者就是一企業機構)裡，要儘量避免自己是個不可採用的人，如果不至少具備一項別人比不上的優點，那在此團體中，有什麼角色能發揮呢？此團體若要裁員時，豈不最早被列出來？

是否有一估計量比其他估計量全都較佳？較佳當然是以 MSE 來評比。除非只有一個可能的 θ，否則便不存在。假設存在一個這種估計量 U，則任取一可能的 θ，不妨以 θ_0 稱之，再取 $V=\theta_0$。則因

$$\mathrm{Var}(V)=\mathrm{Var}(\theta_0)=0，$$

且

$$b(\theta_0,V)=b(\theta_0,\theta_0)=E(\theta_0)-\theta_0=0，$$

因而 $R(\theta_0,V)=0$。故若 U 要比 V 為佳，則須滿足 $R(\theta_0,U)=0$。但 θ_0 不過是任一可能的 θ，因而須有對每一可能的 θ，皆有 $R(\theta,U)=0$。但除了一些退化的情況，此乃不可能。由於不存在一永遠的第一名，在比 MSE 之大小外，我們得再加上其他評比的準則，也藉此排除一些不合理的估計量，如前述 $V=\theta_0$。在較小的估計量之集合中，尋找最佳(指 MSE 最小)估計量，便較可能找到。底下給一例。

設 X_1，X_2，…，X_n，為一組由 $U[0,\theta]$分佈所產生之隨機樣本，其中欲估計的參數 $\theta>0$。對 $n\geq 1$，考慮 X_1，X_2，…，X_n之順序統計量(order statistics)$X_{(1)}$，$X_{(2)}$，…，$X_{(n)}$，即將 X_1，X_2，…，X_n，按小至大排列。令

$U_1=X_{(n)}$，

$U_2=((n+1)/n)X_{(n)}$，

$U_3=X_{(1)}+X_{(n)}$，

$U_4=(n+1)X_{(1)}$，

$U_5=2S_n/n$，

其中如前 $S_n=X_1+X_2+\cdots+X_n$。上述 5 個統計量，皆可用來當做 θ 之估計量，底下我們來比較其 MSE。

計算過程略去，5 個估計量之 MSE 分別為

$R(\theta,U_1)=2\theta^2/((n+1)(n+2))$，

$R(\theta,U_2)=\theta^2/(n(n+2))$，

$R(\theta,U_3)=2\theta^2/((n+1)(n+2))$，

$R(\theta,U_4)=n\theta^2/(n+2)$，

$R(\theta,U_5)=\theta^2/(3n)$。

按小至大之排序如下，這是對每一可能的 θ 皆成立的：

$$R(\theta,U_2)<R(\theta,U_1)=R(\theta,U_3)<R(\theta,U_5)<R(\theta,U_4)，n\geq 2。$$

至於 $n=1$ 時，5 個估計量之 MSE 皆相等，即有

$$R(\theta,U_1)=R(\theta,U_2)=R(\theta,U_3)=R(\theta,U_4)=R(\theta,U_5)，n=1。$$

因而 U_1、U_3、U_4，及 U_5，皆為不可採用的。當然這並不保證 U_2 為可採用的。

再度 X_1，X_2，⋯，X_n，在區間[0,θ]均勻分佈，按小至大

排出來為 $X_{(1)}$，$X_{(2)}$，…，$X_{(n)}$，這 n 個點將$[0,\theta]$分割成 $n+1$ 個子區間。均勻分佈的關係，不難看出，平均來說每一子區間的長度皆為

$$\theta/(n+1)。$$

$X_{(n)}$是 n 個點中最接近 θ 的一個，這是何以會想到估計量 U_1。但畢竟 $X_{(n)}$比 θ 小，以 $X_{(n)}$來估計 θ，必然是低估，即 U_1 為偏差的。調整一下，乘上$(n+1)/n$ 後，得到 U_2。至少 U_2 滿足 $E(U_2)=\theta$，為不偏的。而也可預期 U_2 必較 U_1 為佳。U_3 的產生，也是因 U_1 比 θ 小，故加上 $X_{(1)}$，彌補一下，至少讓 U_3 因而成為不偏的。U_1 為偏差的，但將 U_1 加上一隨機的量所得的 U_3，雖成為不偏，MSE 是否會變大或變小，可就難說了。事實上，二者之 MSE 仍相等，這倒是一有趣的現象。由於 $X_{(1)}$離 θ 較遠，乘上$(n+1)$後，雖使 U_4 成為不偏的，但變異數顯然會較大，不像 $X_{(n)}$已接近 θ 了，經由乘上一個接近 1 的因子$(n+1)/n$，不太會產生過大的變異。果然，U_4 的 MSE 是最大的。S_n/n 是樣本平均，$E(S_n/n)=\theta/2$。故 U_5 為不偏的。以樣本平均來調整的估計量，應不如以 $X_{(n)}$來調整的估計量，但優於以 $X_{(1)}$來調整的估計量，而也的確有 $R(\theta,U_2)<R(\theta,U_5)<R(\theta,U_4)$，$n\geq 2$。

欲估計參數 θ，由(1)及(2)式知，一估計量 T 若為不偏的，則 $b(\theta,T)=0$，且此時

(3) $R(\theta,T)=\mathrm{Var}(T)$。

我們常可在所有不偏估計量中，找到一較所有其他估計量至少不差的估計量。即常會存在一不偏估計量 T^*，使得對任一其他的不偏估計量 S，及任一可能的 θ，皆有

$$R(\theta,T^*)=\mathrm{Var}(T^*)\leq\mathrm{Var}(S)=R(\theta,S)。$$

這種 T^* 便稱為“一致最小變異不偏估計量”(uniformly minimum variance unbiased estimator，簡稱 UMVUE)。在此所謂“一致”，是指對一可能的 θ，T^* 之變異數皆最小。

既不偏，變異數又最小，其涵義為何？仍以射箭為例。平均沒有偏差，且射在靶上的點都很接近，那不就是射出的箭都集中在靶心附近？射箭技術顯然高明。一般而言，UMVUE 是統計裡認為不錯的，也有一些有效的定理來協助尋找。但有時這樣的估計量，卻顯得極不合理。譬如說，在估計一絕不會是負的參數 θ，找到的一致最小變異不偏估計量，取值卻可能為負，例子可參考一般數理統計的教科書。另外，尚有幾個評比的準則，此處僅是初步的介紹，我們就此打住。

34 信賴區間

我們已介紹了點估計，即以一個數值，來估計一未知的量。但點估計通常不會準確，甚至在很多情況下，估計正確的機率為 0，區間估計因應而生。某日小明考完數學，媽媽問他能考幾分？這相當於要小明估計。小明較少會回答諸如 85 分這種單一值，比較可能說出如 8、90 這類。這便是區間估計。媽媽懷疑小明過度樂觀，再問有幾成把握？9 成！小明充滿信心地回答。幾成把握即表機率有多大。小明便是說，他的分數，會介於 80 至 90 分間的機率為 0.9。對一隨機現象，其中未知參數之點估計，並不易命中真實值。於是除了點估計外，遂發展出區間估計。即以一區間來估計某參數(表未知的量)，並附上該參數會落在此區間之機率。其中的區間，便稱為信賴區間，至於伴隨的機率(如前述小明說的 9 成)，則稱為信心水準。在小明的成績估計裡，信賴區間即 [80,90]，信心水準則為 90%。

曾有人注意到，經濟上的預測，習於只給一明確的值，也就是只採點估計。可能認為給一區間，顯示預測者缺乏信心。而若連自己都沒信心，做出來的預測，豈值得信賴？因而學者或官員，對未來經濟的預測，若以區間估計呈現，往往會被質疑是因信心不足、專業性不夠，或擔心達不到目標而挨罵等。醫生這一行，處境常也類似。有人就是希望醫生給一個斬釘截鐵的答案，如手術成功的機率有幾成？若答 5

至 7 成，會被懷疑這樣的模稜兩可，顯然醫術不高明。看起來，是有一些人，並不以為信賴區間值得信賴。至於民調，於公佈調查結果時，除估計值外，亦會給出正負誤差，那就是給估計參數的信賴區間了。何以民調預測這一行，能接受信賴區間？應是大眾已了解民意向來如流水，知道人的想法就是會改變，以區間來表示預測結果，乃是科學的。

信賴區間唯一嗎？這可能是立即會想到的問題。在估計一參數時，若決定採取區間估計，常是先給定信心水準，然後求出一區間，使該區間涵蓋參數之機率，如事先所設定的信心水準。信心水準一般以百分比表示，95%是常取的信心水準，當然 99%、90%，或其他百分比都可以。對一固定的信心水準，信賴區間往往並不唯一。那如何選擇？理想狀況當然是取區間長度最短的。區間長度愈短，表估計愈精準，雖有時最短區間並不易找得到。其次，為什麼不說機率，而說信心水準？也就是 95%究竟是什麼，代表機率嗎？但又怎能是機率？要知取樣前信賴區間是一隨機區間，說有 95%的機率會包含待估計的參數，這點較無疑義。但取樣後則得一常數區間，或者說固定的區間，此區間要嘛包含參數，要嘛不包含，如何能說包含的機率是 95%？甚至，若多次重新取樣，將得到多個不同的信賴區間，難道這些不同的區間，每一皆有一樣的機率，涵蓋同一參數？本來人們較不常去想機率究竟是什麼，但接觸信賴區間後，不少人卻連機率的涵義，也開始感到疑惑了。事實上，如我們之前曾提過的，只要是未知便能談機率。正如教室講台抽屜有 3 個蘋果及 2 個

梨，老師拿了一個放背後，雖究竟是蘋果或梨乃確定，仍可問學生是蘋果的機率為何？又如考前老師早已出好題了，學生仍可猜題，說這題 8 成會考，那題會出現的機率連 0.01 都不到等。再如單位裡某人氣相當旺的女同事，懷胎 9 月即將臨盆，雖肚子裡的寶寶性別早已確定，但好事者仍屢在猜測她生男或生女的機率。同理在估計銅板出現正面的機率 p 時，由於 p 未知，因而取樣後所得的一個固定信賴區間，是可以對它談機率，即會包含 p 的機會為 0.9，或 0.95 等。其實就是用主觀機率，將信心水準，視為該區間包含 p 的機率，而這乃合情合理。

假設某廠牌電池的壽命有常態分佈 $N(\mu,\sigma^2)$。由於製程穩定，變異數 σ^2 可設為已知。現欲估計期望值 μ。取 n 個電池來測試，得 X_1，X_2，…，X_n，分別表各電池之壽命，可假設為相互獨立且以 $N(\mu,\sigma^2)$為共同分佈，因而 S_n/n 有 $N(\mu,\sigma^2/n)$ 分佈，其中 $S_n=X_1+X_2+\cdots+X_n$，$n\geq 1$，為樣本和，而 S_n/n 當然便是樣本平均。由於$(S_n/n-\mu)/(\sigma/n^{1/2})$有標準常態分佈 $N(0,1)$。故

$$P(-1.96\sigma/n^{1/2}\leq(S_n/n-\mu)\leq 1.96\sigma/n^{1/2})\approx 0.95，$$

其中 1.96 是由 95%的信心水準而來。上式導致

$$P(S_n/n-1.96\sigma/n^{1/2}\leq\mu\leq S_n/n+1.96\sigma/n^{1/2})\approx 0.95。$$

由是得 μ 之一(近似的)95%信賴區間為

$$[S_n/n-1.96\sigma/n^{1/2}, S_n/n+1.96\sigma/n^{1/2}]。$$

上述區間之長度為 $3.92\sigma/n^{1/2}$，可看出 n 愈大時，區間長度愈短，且 $n\to\infty$ 時，$3.92\sigma/n^{1/2}\to 0$。付出代價愈高(樣本數愈大)，可得愈精準的估計。由此也可看出，至少對不同的樣本數，信賴區間並不唯一。

我們給的信賴區間乃對稱於 μ，可否不對稱於 μ？可以，但對常態分佈，只要同一 n，不對稱的信賴區間，便非最短。如因

$$P(-1.8814\sigma/n^{1/2}\leq(S_n/n-\mu)\leq 2.054\sigma/n^{1/2})\approx 0.95。$$

由是得 μ 之一(近似的)95%信賴區間為

$$[S_n/n-2.054\sigma/n^{1/2}, S_n/n+1.8814\sigma/n^{1/2}]。$$

上述區間之長度為 $3.9354\sigma/n^{1/2}$，比原本的 $3.92\sigma/n^{1/2}$ 長些。事實上，在本例中，有無限多個 μ 之 95%信賴區間，而對稱的那個最短。

當 σ 未知時，對 μ 的估計，有不同的作法；而也能估計 σ^2，也是分 μ 已知或未知，有不同的信賴區間；又，亦可同時估計 (μ,σ^2) 之信賴區間。另外，對不同分佈的參數，也可給出參數估計之信賴區間，在此均略過不提。統計裡的分佈很多，即使常用的都有十餘種，每種的參數從 1 到 3 個都有，甚至更多。由於隨機變數和的分佈，多半無簡潔的形式，故常得藉助中央極限定理，求出近似的信賴區間。因而大學裡有關信賴區間的章節，琳琅滿目，涉及各種分佈及各種分佈的估計，有夠多的題材可討論。另一方面，之前已學過點估

計，學生不難輕易接受區間估計，因不過是點估計的擴展。至於中央極限定理，也在之前便已學過了。因而在大學統計裡，信賴區間並非一個會太讓學生感到深奧的題材。當然可能會覺得，求一個又一個不同分佈及不同參數之信賴區間，有點繁瑣。

35 高中數學裡的信賴區間

自“九五課綱”起，高中數學裡引進信賴區間，直到“108 課綱”才取消，解除了十餘年來，高中數學課程裡的困擾。底下藉兩道考題，以了解信賴區間何以不適合放在高中數學裡。先看 98 學年，學測數學裡的一道多選題：

> 某廠商委託民調機構在甲、乙兩地調查聽過某項產品的居民佔當地居民之百分比(以下簡稱為“知名度”)。結果如下：在 95%信心水準之下，該產品在甲、乙兩地的知名度之信賴區間分別為[0.50,0.58]、[0.08,0.16]。試問下列那些選項是正確的？
>
> (a) 甲地本次的受訪者中，54%的人聽過該產品。
>
> (b) 此次民調在乙地的受訪人數少於在甲地的受訪人數。
>
> (c) 此次調查結果可解讀為：甲地全體居民中有一半以上的人聽過該產品的機率大於 95%。
>
> (d) 若在乙地以同樣方式進行多次民調，所得知名度有 95%的機會落在區間[0.08,0.16]。
>
> (e) 經密集廣告宣傳後，在乙地再次進行民調，並增加受訪人數達原人數的四倍，則在 95%信心水準之下該產品的知名度之信賴區間寬度會減半(即 0.04)。

“大考中心”公佈的答案為(a)、(b)。

這可能是大學統計課程裡，極罕見的題型。在大學教統計的教授，相信有不少看到這種考題時，應會傻眼，無從答起。對這類民調，高中數學課本通常給出型式如下的信賴區間：

(1) $[(S_n/n-1.96(S_n/n(1-S_n/n)/n)^{1/2}, S_n/n+1.96(S_n/n(1-S_n/n)/n)^{1/2}]$，

其中如前 $S_n=X_1+X_2+\cdots+X_n$，$n\geq 1$，而 X_1，X_2，…，X_n 分別表分別表依序調查所得之結果，而 $X_i=1$，或 0，就依第 i 個被訪問的居民，是否聽過該項產品，$i=1$，…，n。S_n/n 表調查後之知名度，即聽過該項產品的居民之百分比。又，1.96 是由 95%的信心水準而來。可看出(1)式中的區間，有 $S_n/n\pm 1.96(S_n/n\ (1-S_n/n)/n)^{1/2}$ 之型式，且區間中心為 S_n/n，區間半徑為 $1.96(S_n/n(1-S_n/n)/n)^{1/2}$。

將所給甲地的信賴區間[0.50,0.58]與(1)式比較，得 $S_n/n=0.54$。故選項(a)正確。甲乙兩地知名度之信賴區間，其中心 S_n/n 分別為 0.54 及 0.12，但區間半徑皆為 0.04。因

$$0.54\times(1-0.54)=0.2484>0.1056=0.12\times(1-0.12)，$$

顯然乙地的 n 較小。選項(b)為正確，也就出來了。看到這裡，不禁令人好奇，這比較像道簡易的數學題目吧！豈有絲毫統計的味道？

選項(c)不在答案裡，此值得斟酌。在統計裡，依理念之

不同，甚至就只是主觀，可有各種推論，及各種解讀。被視為一意孤行或見解偏頗者，還可能自以為擇善固執或獨具慧眼呢！數學裡就不行，有一套邏輯，不可自由推論、隨意解讀。選項(c)既然問此次調查結果可否“解讀”為…，光憑問句中的“解讀”二字，此選項就可判定為正確。何況題目裡說，“在 95%信心水準之下，該產品在甲地的知名度之信賴區間為[0.50,0.58]”。如前指出，採主觀機率，先將題意解讀為知名度會有 95%的機率，落在區間為[0.50,0.58]。由此得知名度有大於 95%的機率是一半以上，因知名度落在[0.50,0.58]，都已有 95%的機率了。再繼續解讀為“甲地全體居民中，有一半以上的人聽過該產品的機率大於 95%”，不正是依民調結果，一連串合理的推論嗎？此正如某機構公佈新近完成之台灣各縣市長的民調，其中某市長之支持度為 32%。有人宣稱，該市長的支持度岌岌可危，比上次調查少了 11%。但有人卻不以為然，因另一常被拿來相比的市長，其支持度也不過 35%，兩人支持度差不了太多。又如對於美國軍事人員在台灣，有人說是“部隊駐台”，有人說是“外國兵力進駐台灣”，但也有人說不過是“交流協訓”，不必大驚小怪。上述幾種不同的解讀，很難說那一正確。另外，二次世界大戰期間，歐洲有將近 6 百萬猶太人被殺害，戰後在世界各地，蓋了不少被害猶太人紀念碑。但亦有些猶太人，對某些紀念碑的設計提出質疑，“我們的傷痛，豈能讓人隨意解讀？”凡此種種，皆說明所謂“解讀”，是沒什麼一致標準的，甚至還常淪於各說各話。因而學測裡出現這類不易有定論，以“解讀”發問的選項，並不恰當。

再來看 99 學年學測數學的一道多選題：

想要了解台灣的公民對某議題支持的程度所作的抽樣調查，依性別區分，所得結果如下表：

	女性公民	男性公民
贊成此議題的比例 $\hat{p}$	0.52	0.59
$\hat{p}$ 的標準差 $\sqrt{\hat{p}(1-\hat{p})/n}$	0.02	0.04

請問從此次抽樣結果可以得到下列哪些推論？

(a) 全台灣男性公民贊成此議題的比例大於女性公民贊成此議題的比例。

(b) 在 95%的信心水準之下，全台灣女性公民贊成此議題之比例的信賴區間為[0.48,0.56](計算到小數點後第二位，以下四捨五入)。

(c) 此次抽樣的女性公民數少於男性公民數。

(d) 如果不區分性別，此次抽樣贊成此議題的比例 $\hat{p}$ 介於 0.52 與 0.59 之間。

(e) 如果不區分性別，此次抽樣 $\hat{p}$ 的標準差 $\sqrt{\hat{p}(1-\hat{p})/n}$ 介於 0.02 與 0.04 之間。

“大考中心”公佈的答案為(b)、(d)。

選項(a)，如前所述，統計考題裡並不適合問可以得什麼“推論”。不過由所給答案，我們來推敲命題者的用意。調查結果，男性及女性之贊成比例，其 95%信賴區間，分別約為[0.51,0.67]，[0.48,0.56]。由於兩區間有重疊，猜想這可能是(a)不被命題者視為正確選項的原因。但比較二未知量的大小，並不一定得依靠信賴區間。若採點估計，因 0.59>0.52，由此得到(a)之推論不行嗎？即使採信賴區間來相比，非得取 95%？採 68%不行嗎？而若採 68%，則男女贊成比例之信賴區間，便各約為[0.55,0.63]及[0.50,0.54]，不再重疊了。在統計裡，可有各種推論法，且常無那一推論法永遠最佳，只能依不同的標準評比。至於選項(b)、(c)、(d)，及(e)，乃考是否熟悉課本上所給信賴區間的公式。(c)、(d)，及(e)，甚至根本是在考數學。

在本人所寫“機率統計考題探討”(2012)一文裡，曾將 98 至 101 學年，學測數學，及指考數學甲、數學乙，4 年間共 12 份試題，挑出機率與統計方面，值得商榷的題目來討論。在該文的結論裡寫著：

> 看來信賴區間快沒題目可考了。…。現今大學入學考，不論學測或指考的數學科，機率統計的考題，有時像在考三民主義，思想要很制式，才易得高分。

本來估計的教學，就宜以估計為主要目的。對想以一區間來估計一參數的教材，便該常在求出信賴區間。大學的統

計課程裡，在信賴區間那章，有各種分佈，及期望值、標準差，還有它們的函數等，且各分佈不乏有不只一個參數可估計。可於不同的情況下，探討如何求得參數的信賴區間。由於包含的內容不少，因此進度慢不下來，教師快速講授，學生匆忙學習，甚至囫圇吞棗，因而根本無暇去思考信賴區間的內涵，如信心水準 95%，其中 95%的意義為何？這可能是何以多年來，大學信賴區間的教學，少見困擾的主因。並非修習學生個個皆能融會貫通，而是問題不太浮現。但在高中數學裡，引進信賴區間唯一的目的，便是為了給出民調時，對某議題估計支持度的信賴區間。既然是民調，不論那一議題，對每位選民的調查結果，不外支持或不支持，這涉及 $Ber(p)$分佈，其中參數 p 就是擬估計的支持度。但由於取樣不放回，對 n 個人調查後，所得之 n 個樣本並不獨立，因而無法利用中央極限定理來近似。給一即使大學數學系畢業生，多半仍一知半解的定理，且定理中已指出樣本須獨立且有相同分佈才適用，如今此定理之唯一且立即的應用，是從事民調時，所得那組不獨立的樣本，條件不滿足還引用該定理，師生怎能不充滿疑惑，感到混亂不已？更不要說其中涉及的機率涵義，絕非三言兩語可講清楚。信賴區間的難度，被將其引進“九五課綱”的學者，嚴重低估了。

就只針對不獨立的 n 個 $Ber(p)$分佈隨機變之和，要求參數 p 的信賴區間、以常態分佈來近似，且信心水準僅設定為 95%。內容如此少，因而求信賴區間的題目，便相當缺乏變化，加上計算又容易，在教學目的幾乎可說僅是考試的高

中，如何命題呢？於是倒過來，先給出信賴區間，然後來“解剖”該區間，前述那些類型的考題遂產生了。即使在大學教了多年統計學的教授，面對這樣的多選題，都會傻眼。他們何時出過如此型式的考題？學測或指考裡的信賴區間考題，曾有大學統計學教授硬著頭皮去做答，卻在 5 個選項中，錯了好幾個。雖然這樣，但翻來覆去，畢竟就只有那幾套題型，久後便被高中師生破解，知道命題者想要的答案。“思想正確”才易答對，那不就像在考三民主義嗎？也因如此，我們才認為快沒題目可考了。事實上，從 102 至 110 學年，9 年間的學測及指考，除了 102 學年指考的數學乙，有一道多選題的信賴區間外，便再無信賴區間的考題了。到這個地步，信賴區間是可移出高中了。而果然。2019 年開始實施的“108 課綱”，就取消信賴區間了。

除了中學引進信賴區間不易講清楚外，因信心水準的引進，使得師生連對機率的涵義都感到疑惑了，再加上考題往往會將學生帶進統計的死胡同，無法讓學生吸收到有意義的統計概念，制訂“108 課綱”時，將信賴區間自高中數學移除，是正確的。信賴區間誤進高中數學十餘年，終於灰頭土臉地離開。

36 風險

嗜賭普遍被認為是不好的習性，不少嗜賭者後來傾家蕩產、眾叛親離。勸人戒賭者，常會說十賭九輸。現假設有某好賭的 A 君不服氣，提出如下一簡單的賭局：每回不論下多少賭注，輸了賭注被收走，贏了則除拿回賭注亦獲賠相同金額。並未限定這是一公正賽局，但 A 君給出如下必勝策略：第 1 次賭 a 元，$a>0$，贏了就停止不賭，輸了第 2 次賭 $2a$ 元。再度，贏了就停止不賭，輸了第 3 次賭 2^2a 元，餘類推。我們來看此是否真為必勝策略。

假設賭到第 k 次結束後停止，則第 k 次贏 $2^{k-1}a$ 元，而前 $(k-1)$次共輸

$$(1+2+\cdots+2^{k-2})a \text{ 元}=(2^{k-1}-1)a \text{ 元。}$$

故賭局停止時，A 君淨贏 a 元。此策略看起來是能必勝，但真能這樣做嗎？底下來分析。

設每次 A 君贏的機率為 p，即輸的機率為 $1-p$，$0<p<1$。若 $p\neq 0.5$，此便非一公正的賽局。但 A 君信心滿滿，認為不論 p 多小，他贏定了。現設至停止時，A 君共賭 X 次，則

$$P(X=k)=p(1-p)^{k-1}\text{，}k\geq 1\text{。}$$

即 X 有參數為 p 之幾何分佈。因 $E(X)=1/p$，若 $p=0.5$，賭局平均 2 次結束。即使 $p=0.1$，賭局平均 10 次結束，賭的次數

總是有限的，且除非 p 很小，否則通常並不必賭太多次。但由

$$P(X \geq k)=(1-p)^{k-1}，k \geq 1，$$

雖平均有限次便能結束賭局，但所需賭的次數，卻可能(機率為正)無止盡的大。事實上，若 A 君的口袋夠深，通常的確能淨贏 a 元，快樂地離去。但當連輸 k-1 次，且 k 較大，導致 A 君擁有的資金已不足 $2^{k-1}a$ 元，則第 k 次他便無法下注了。除非賭場能讓他欠錢。但賭場怎會搬磚頭砸自己的腳？所以這並非一真的必勝策略。不過，由此我們得到啟示，對資金夠雄厚者，如果野心不太大(即 a 不大)，投資時不冒太大的風險(如避開 p 太小者)，則便較易(機率較大)達到目標。而資金不夠雄厚卻野心過大者，通常便不易達到目標。

再看一例子。設 B 君參加前述賭局，每回賭注多寡不拘，且每回他獲勝之機率為 0.4，輸之機率為 0.6。設 B 君有 n 個籌碼，我們來比較二策略。策略(一)乃一次 n 個籌碼全下注，且只賭 1 次。令 S_n 表賭局結束後，B 君之淨所得。則可求出

$$E(S_n)=-0.2n，\mathrm{Var}(S_n)=0.96n^2。$$

策略(二)乃每次賭 1 籌碼，共賭 n 次，令 T_n 表賭局結束後，B 君之淨所得。則可求出

$$E(T_n)=-0.2n，\mathrm{Var}(T_n)=0.96n。$$

在賭局對 B 君不利下，策略(一)的孤注一擲，與策略(二)之

分散風險，B 君淨所得之期望值不變，都是-0.2n，但前者之變異數為後者之 n 倍。即在本例中，B 君若採分散風險策略，將使損失之變異數減小。

在上例中，當 n 夠大時，由中央極限定理，

$$(T_n+0.2n)/(0.96n)^{1/2}$$

其分佈近似於 $N(0,1)$。故

(1) $P(T_n<0)\approx P(Z<0.2n/(0.96n)^{1/2})$，

且

(2) $P(|T_n+0.2\mathrm{n})/(0.96n)^{1/2}|\leq 3)\approx P(|Z|\leq 3)$，

其中 Z 有 $N(0,1)$分佈。現取 n=10,000，代入(1)式，得

$$P(T_{10,000}<0)\approx P(Z<2{,}000/(9{,}600)^{1/2})\approx P(Z<20.41)，$$

超過 20 個標準差，不必查表，也不必用計算器，此機率幾乎是 1 了。即在賭局不利下，若採孤注一擲的策略，則 B 君有 0.4 的機率，淨所得為正數 10,000；但若採分散風險策略，則 B 君之淨所得幾乎(機率極接近 1)是負的。次將 n=10,000，代入(2)式，則有

$$P(|(T_{10,000}+2{,}000)/(9{,}600)^{1/2}|\leq 3)\approx P(|Z|\leq 3)\approx 0.9973。$$

即

$$P(-2{,}294<T_{10,000}<-1{,}706)\approx 0.9973。$$

因而不像採策略(一)，有 0.4 的機率，B 君之淨所得為正數10,000，採策略(二)，淨所得有極大機率(0.9973)，介於-2,294至-1,706 間。雖幾乎是穩輸，但不像採策略(一)，有不小的機率(0.6)輸很多(10,000)。此例說明，何以在賭場、商場或戰場上，居劣勢的一方，屢採孤注一擲策略，因這樣仍有贏的機會，雖一旦輸就是損失很大。至於步步為營者，雖損失不會過大，卻沒什麼取勝的機會。

買保險有必要嗎？既說是“險”，便知其中有風險。有人樂於買保險，沒理賠時覺得很好，畢竟健康平安最幸福，而一旦生病住院醫療，獲得理賠時，便覺幸好有投保。保險公司在負擔包含員工薪資等各種行政費用，還要獲利，因而對投保者而言，直觀上，其希望淨所得必為負值，即這是一非公正的賽局。但之所以願意投保，並非為了賺錢，乃基於每年的保費自己負擔得起，而世上災禍難料，若因醫療需要，得有大筆支出，此時便有保險做後盾了。另一方面，保險公司也並非必然獲利，它也是有風險的。眾所周知，保險公司乃向投保人收取保費，以支付經營的各種開銷。保費要仰賴精算師(actuary)細算，利用死亡、傷害及意外等各種數據，以制定保單。但即使經過精算，也難免會遇到很多人同時要求理賠的情況，當然那可能百年才一次。如 2001 年 9 月 11 日，發生在美國包括紐約市在內數地的九一一襲擊事件(September 11 attacks)，因該事件而死亡或失蹤的總人數，至少有 2,996 人。又，並非只有恐怖攻擊，才會導致大量死亡，也有人謀不臧的。如因隨意更動建築設計，導致 1995 年 6

月 29 日，南韓三豐百貨店倒塌事故(Sampoong Department Store Collapse)，造成 502 人死亡，937 人受傷。突如其來的意外，是否會賠不出來？也就是保險公司本身保險嗎？底下以一簡單的例子來說明，保險公司若未適當規畫避險的方式，便絕非穩賺不賠。

根據衛生福利部之統計，2020 年，台灣因“事故傷害”而死亡的人數為 6,767 人。至 2020 年底，台灣的人口約 2,357 萬，故 2020 年的意外死亡率不妨以

$$6{,}767/23{,}570{,}000 \approx 2.87 \times 10^{-4}$$

為估計值。現假設 C 君成立一家保險公司，只承辦意外事故導致身故才理賠。保費每案一律每年 4 千，理賠金額則為 5 百萬，單位皆為台幣的元。又設每一投保案之平均行政支出費用為 1,100。以 Y 表某一投保者在 2020 年之淨所得，則

$$E(Y) \approx 2.87 \times 10^{-4} \times 5 \times 10^{6} - 4{,}000 = 1{,}435 - 4{,}000 = -2{,}565$$

為一負值。但該投保者並不太在意，因他是為了一旦遭遇意外身故，可留給家人一個保障，投保可不是為了被理賠。至於對 C 君而言，以 W 表 2020 年，對某一投保者，所淨獲之利益。則

$$E(W) \approx 4{,}000 - 2.87 \times 10^{-4} \times 5 \times 10^{6} - 1{,}100 = 1{,}465。$$

對每一投保案，C 君平均淨獲利益約 1,465，雖不如投保者以為的約有 2,565 那麼高，但每案之獲利超過 35%，比很多行

業好多了。那 C 君是否得承擔風險呢？

假設只有 1 千人投保，則這 1 千人皆未意外身故的機率約為$(1-2.87\times10^{-4})^{1,000}$，因而其中至少有 1 人意外身故的機率為

$$1-(1-2.87\times10^{-4})^{1,000}\approx1-0.750=0.250。$$

約 1/4 的機率，並不算小。保費扣除行政支出後之總額為

$$(4,000-1,100)\times1,000=2,900,000，$$

因此只要理賠 1 人，便虧了 210 萬，在經營初期，C 君可說常處於心驚膽跳的狀態。若保戶中，有一群相識者常呼朋引伴來投保，而某回他們搭同一輛遊覽車出遊，卻不幸翻車遇難身亡，C 君若資金不夠雄厚，這時將賠慘了。但由中央極限定理可求出，只要保戶夠多，如達到 10 萬，則 C 君會虧的機率便很小；若保戶夠達到百萬，則 C 君會虧的機率，便將微乎其微。但要能撐到保戶達到百萬，資金當然得相當雄厚。

再看一開始那一賭局。設 D 君每次投注 1 元，輸了賭注被收走，贏了除拿回賭注亦獲賠 1 元。又設每次 D 君贏的機率為 p，即輸之機率為 $1-p$，$0<p<1$。令 U_n 表賭 n 次後，D 君之淨所得。則易求出

$$E(U_n)=n(2p-1)，\mathrm{Var}(U_n)=4np(1-p)。$$

當 n 夠大時，由中央極限定理，

$$P(U_n\geq 1)=P((U_n-n(2p-1))/(4np(1-p))^{1/2}\geq(1-n(2p-1))/(4np(1-p))^{1/2})$$

$$\approx 1-\Phi(((n+1)/2-np)/(np(1-p))^{1/2})，$$

其中 $\Phi(x)=P(Z\leq x)$，$x\in R$，且 Z 有 $N(0,1)$分佈。現假設 D 君所賭為歐式輪盤(European roulette)，即有 1 至 36 的數字，奇數為紅色，偶數為黑色，再加上一綠色的 0。押紅黑中的某色，0 出現算莊家贏，即 $p=18/37$，賠率仍設為 1 賠 1。對一些不同的 n，我們給出 $P(U_n\geq 1)$之近似值如下：

$n=10^2$，$P(U_n\geq 1)\approx 0.3555$，

$n=10^3$，$P(U_n\geq 1)\approx 0.1876$，

$n=10^4$，$P(U_n\geq 1)\approx 3.327\times 10^{-3}$，

$n=10^5$，$P(U_n\geq 1)\approx 5.997\times 10^{-18}$。

即設 D 君之欲望不高，賭 n 次，只要有贏(即 $U_n\geq 1$)即可。若只賭 1 次，$P(U_1\geq 1)=18/37\approx 0.4865$，D 君的心願尚不難達成。但隨著 n 之增大，D 君將愈來愈難如願。要知對大部分的賭客，一旦上了賭桌，便不易下來，所以他們的 n 是很大的。因而在此情況下，賭場並不太擔心風險問題。

在“三國演義”第九十五回裡，蜀國承相諸葛亮(181−234)在第一次北伐(即一出祁山)時，魏國派出驃騎大將軍司馬懿(179−251)前去對抗。司馬懿對先鋒張郃(生年不詳−231)論斷諸葛亮，“諸葛亮平生謹慎，未敢造次行事。若是吾用兵，先從子午谷逕取長安，早得多時矣。他非無謀，但

怕有失，不肯弄險。”子午谷南北縱向，長約 330 公里，北起陝西省長安縣西南秦嶺山中，南至石泉縣。“不肯弄險”在此處，被司馬懿視為諸葛亮之一缺點。但在同一回裡，守在西城的諸葛亮，於接獲飛馬來報，說司馬懿引大軍 15 萬望西城蜂擁而來。由於身邊將領已全派遣出去，且城中僅餘一班文官及二千五百軍，只好擺出空城計。從遠處望到此情況的司馬懿，立即下令退兵。其次子司馬昭(211−265)說，“莫非諸葛亮無軍，故作此態？父親何故便退兵？”司馬懿回答，“亮平生謹慎，不曾弄險。今大開城門，必有埋伏。我兵若進，中其計也。汝輩豈知？宜速退。”待魏軍遠去後，餘悸猶存的眾文官，立即問諸葛亮，統率 15 萬精兵的司馬懿，乃魏之名將，怎麼從遠處，看到承相坐在城樓焚香操琴、悠閒自得，便趕緊撤軍？諸葛亮答道，“此人料吾生平謹慎，必不弄險；見如此模樣，疑有伏兵，所以退去。吾非行險，蓋因不得已而用之。”原來弄險、不弄險，有時恰當，有時不恰當。

我們雖藉數例，以機率的角度來分析風險，但在實務上，若只賭 1、2 次，有些人便不見得認同，僅憑機率的紙上談兵，為恰當的決策。反而習於採弄險或不弄險式地隨機應變，且以成敗論英雄。雖然如此，如同在介紹主觀機率時所說，即使採隨機應變，仍可先評估一下機率。“三國演義”第九十三回，當魏國軍師王朗(153−228)，前來企圖以話術逼迫諸葛亮不戰而退兵。諸葛亮心想，“王朗必下說詞，吾當隨機應之。”結果一席話下來，反讓王朗聽後，“氣滿胸膛，

大叫一聲，撞死於馬下。”“三國演義”裡，將諸葛亮描述成極擅長隨機應變。但在陳壽(233–297)著的“三國志”之“諸葛亮傳”裡，於評論諸葛亮時，說他“可謂識治之良才，管、蕭之亞匹矣。然連年動衆，未能成功，蓋應變將略，非其所長歟！”小說裡或許會過於渲染，陳壽則認為諸葛亮的應變能力有所不足，由於是正史，說不定較接近真實。但無論如何，不論在小說及正史中，均顯示面對風險時，隨機應變的能力是很重要的。

37 從迴歸效應談起

鐵的紀律或愛的教育，何者對提高成績較有幫助？這可能不容易回答。有位小學低年級教師，一向嚴責考試成績不佳的學生，但對考得好的學生，不要說獎勵了，連口頭讚美也吝惜。多年的教學經驗告訴他，本來考得較差的學生，經處罰後，往往下次會進步；至於原先考在前面者，常有不少下次成績退步了。這位教師相信，若當初還獎勵或讚美考得較好者，豈不是會讓他們退步更多？處罰考差者，且不讓考得好者鬆懈，最後受惠的是學生，他認為這才是真為學生好。他以此經驗傳授給新進教師，只是有細心的新教師注意到，那些秉持愛的教育，對學生一向鼓勵有加的教師，班上學生成績，也屢有上次考得好，下次下降；上次考得差，下次上升的現象。啊！年輕教師猛然想起，大學時修過統計學，其中曾提到一著名的迴歸效應(regression effect)。原來對低年級學生，那會了解什麼少壯不努力，老大徒傷悲？那會了解什麼學如逆水行舟不進則退？考試到了，就是去應考，並未特別在意。因而高分下降或低分上升，乃稀鬆正常，跟教師對待學生是否嚴厲，並無太大關係。

為了描述父親身高與兒子身高兩變數間的關係，19 世紀時，英國學者高爾頓(Sir Francis Galton，1822−1911)，引進了迴歸的概念，由此衍生出迴歸分析此一現今常用的統計方法。高爾頓先注意到，個子較高的父親，兒子亦有較高的傾

向，至於較矮的父親，兒子也有較矮的傾向。這當然不是什麼了不起的發現，僅是眾所皆知的遺傳之影響而已。但隨後高爾頓又觀察到，較高的父親，兒子往往比父親稍矮些，而較矮的父親，兒子卻屢比父親略高些。高爾頓稱此現象為“向平均迴歸”(regression toward the mean)，也就是靠向平均。其中“迴歸”(regression)一詞，即有後退或退化的意思。後來便稱此為“迴歸效應”。在高爾頓所獲得的 1,078 對父子的身高數據中，父親的平均身高約為 68 吋(=172.72 公分)，兒子則約為 69 吋(=175.26 公分)。既然兒子比父親平均約高 1 吋，因此直觀上，對一個 72 吋(=182.88 公分)比平均高的父親，會估計其兒子身高 73 吋，而對一個 64 吋(=162.56 公分)比平均矮的父親，會估計其兒子身高 65 吋。實際上，分別約為 71 吋及 67 吋，前者較預期矮了 2 吋，後者則較預期高了 2 吋，皆向兒子身高的平均值 69 吋靠近。雖然高父親還是有高兒子，矮父親還是有矮兒子，但平均而言，下一代的平均身高拉近了。

是上天濟弱扶傾，有意做一些平衡，才產生迴歸效應嗎？倒也非如此。就仍以父子身高為例，來說明迴歸效應產生的原因。高爾頓的時代，由於營養及生活環境逐漸改善，人們的身高隨之一代一代略增，但也沒增太多，今日則已趨於平穩了。假設某國男子最高的達 235 公分，又假設這些 235 公分高的長人，他們兒子的平均身高沒有比較高，仍為 235 公分。但因不致於全部長人家庭，下一代都一般高，故即使那群兒子的平均身高為仍為 235 公分，則其中必有比 235 公

分還高者，當然也會有比 235 分還矮者。如此一代超越一代，該國的長人身高紀錄，將不斷提高。但人類的身高紀錄，倒也沒有一直持續上升。故身高 235 公分的那群父親，其兒子的平均身高，不要說不會多於 235 公分，比較可能會少於 235 公分。反過來，最矮的那群父親，其兒子的平均身高，也很可能會比其父親輩增加些。

考試成績也不乏迴歸效應。第一次考高分者，除因本身程度較佳之外，運氣大抵不會壞。如此第二次考試，就“較不容易”有比原先更多的好運了。例如，假設某科的滿分為 100，則第一次考 95 分者，往上只有 5 分的成長空間，要提高豈會容易？下降還較可能。況且，考 95 分者，有些實際程度說不定是 90 分，但因好運，不但會的大都寫對，且做對一些向來不太會的題目，因而多出 5 分，達到 95 分。多 5 分已不容易了，下次再考，合理的預期，是不會有更佳的運氣了。而第一次只考 10 分的人，除了程度較差外，運氣大約也不會太好，連猜都大部分猜錯。而既已在谷底，第二次又能怎樣往下掉呢？運氣豈易比不佳更不佳？也不妨這樣想：為什麼人們會擔心一試定終身？因覺程度不見得能在每次考試，都實在地反映出來。考試成績乃真正程度，加上誤差後之呈現。誤差總是難以避免，會有些機運，可能正向也可能負向。即使每次考試成績為獨立，但如果再考一次，便須放棄原有的成績，則這次考不好者，會較樂於再試一次；若這次得到的成績，已大致不少於自己應有，則是否該再試一次，就得長考了。

胡適(1891–1962)在“嘗試歌”裡寫著，“自古成功在嘗試！莫想小試便成功，那有這樣容易事。有時試到千百回，始知前功盡拋棄。”雖胡適鼓勵人們要多嘗試，但了解迴歸效應後，是否一旦居於領先，只要守成就好，免得愈試愈差，而落後者，則可盡量一試再試，因會愈來愈好？

我們曾在"談不確定性”(2013)一文裡說：

> 宇宙的運轉，乃必然性與隨機性，交錯著進行。天體運行及科學的很多領域中，充滿著必然性。中學的平面幾何裡，便有極多必然性的結果。例如，任一直角三角形，兩股平方和等於斜邊平方，此即著名的畢氏定理，…。在隨機世界裡，必然性使人們願意事先好好準備，因花了功夫，總有相當成效，所謂努力就有好收護；隨機性則使人們對未來充滿著不確定與盼望，因此會戒慎恐懼且不輕易放棄。光有必然性的世界，單調而無變化，將令人對未來失去憧憬，少了努力的動機。光有隨機性的世界，一切只靠運氣，凡事都不確定，將令人不想積極認真。想想若自進高中起，全班的成績排序，三年下來都維持不變；或另一極端，每學期名次全由抽籤決定，那將是什麼情景？必然性加上隨機性，使人們在困境時，會期待隨機性來扭轉，在順境時，會期待必然性能常保。永遠有期待，不論這個世界是如何形成的，必然性與隨機性的並存，此結構實在相當巧妙。看過“千鈞一髮”

(Gattaca，1997)嗎？電影裡，每個人只要一出生，便依照基因，被指定可做什麼工作。基因好的裘德洛(Jude Law，1972−)，卻因某意外事件，造成半身不遂，退出太空人的訓練。天下畢竟不是只有必然性，基因再好也阻止不了意外的發生。而一心想當太空人的伊森霍克(Ethan Hawke，1970−)，卻屬於基因不良者，早就被排除當太空人的可能。伊森霍克放棄了嗎？憑努力加上幸運，伊森霍克最終打敗了基因，升上太空。

子女看著父親或母親成就高，想到不易超越，會感到沮喪嗎？要知迴歸效應乃是就整體平均而言，但高個子父親兒子更高、分數已很高下次更高，這類例子一直也不少。機率也是一樣，雖為做決策時之重要依據，但即使整體而言，顯示嘗試不宜，卻仍因人而異，不時有人試圖突破。如某事做成功的機率為 0.01，一般人聽了會打退堂鼓。但就是有人不信邪，覺得自己非一般人，願意去嘗試。在經充分準備後，他成了 1 百人中成功的那一個。所以，迴歸效應可能會被打敗，機率可能會被打敗，幾千年來，人類的很多大重大發現或進展，常是依靠那些不相信平均者，認為自己非正常人(normal people)者。

38 資料探勘(一)

有人認為統計與考古之工作乃類似，都是在挖掘。考古上窮碧落下黃泉，挖掘出各種遺物及遺跡，統計則從浩瀚的資料中，挖掘出有用的資訊。近年來由統計學衍生出資料探勘(亦稱資料採礦，data mining)這領域，這是從挖掘的角度來看統計，如此讓統計學更具有考古學的味道。至於考古學，也有人認為，其工作並不僅是在四處挖掘，大部分的時候，其實是在做資料分析，以重建古代人類的歷史與文化，如此看來，考古學也頗有統計學的味道了。事實上，除了挖掘及資料分析外，統計學與考古學還有一特質相同。

1931 年，“殷墟”(商朝後期王都遺址，位於今日河南省安陽，以殷都區小屯村為中心)第 4 次挖掘時，才剛加入中央研究院歷史語言研究所(1928 年於廣州成立，簡稱史語所，首任所長為傅斯年(1896−1950))考古組(組長為李濟(1896−1979))的郭寶鈞(1893−1971)便曾說，“事實至於遺存，推論敬俟卓識”(即只報導事實，其含意則有待高明)。因考古這門學問，有物而無言，幾千年前的歷史，由所掘出的古物，如何告訴世人，有關此物之來龍去脈？這可非挖掘出來即可得知，要有專家來分析解讀。要知挖掘固然困難，但更難的是推論。統計不也一樣嗎？數據取得不易，而一旦蒐集到，數據並不會說話，是人在說話。對所得的數據，儘可能給出能被接受的推論，才是統計之目的。附帶一提，殷

墟挖掘由李濟領導，是中國歷史上首次重大的挖掘，不但對中國考古學往後的發展影響深遠，亦屬“20 世紀最偉大的挖掘之一”。

推論並不容易，如舉世聞名，1986 年，在四川省廣漢出土的三星堆遺址，兩個內有許多珍寶的器物坑，究竟是屬於陪葬坑或祭祀坑，至今仍無定論。甚至因蜀地無象，而遺址中卻出土象牙及象骨，因而其來源及用途，也一直爭論不休。那已被接受的推論，會被推翻嗎？也是會的。“霸陵”是漢文帝(西元前 203−157 年，西元前 180−157 年在位)的陵寢，之前大部分的學者認為，霸陵位於今日陝西西安東郊白鹿原的東北角，一個當地人稱為“鳳凰嘴”之處。鳳凰嘴附近有許多碑石，立碑者包括清朝皇帝康熙(1654−1722，1661−1722 在位)及雍正(1678−1735，1722−1735 在位)兩位皇帝。遙想“文景之治”，讓人發思古之幽情，畢竟漢文帝在歷史上的名聲不錯。可惜這幽情，可能從元朝起，便發錯對象了。2021 年 12 月 14 日，大陸國家文物局宣布，經過考古確認，位於鳳凰嘴之南約 2.1 公里(仍屬白鹿原)的“江村大墓”，才是真正的漢文帝霸陵所在地。考古之判斷，如同統計之推論，是可能有誤的。

底下給一考古裡的資料分析之例。王懿榮(1845−1900)是清朝翰林，曾任國子監祭酒(相當於皇家大學校長)。他相當博學、酷愛金石，且喜歡收藏古董。居官不過偶然之事，他一生最大的貢獻，乃是發現甲骨文，被推崇為“甲骨文之父”。由於本身略通醫術，他向來會查看家人從藥店抓回來

的藥材，以了解自己究竟吃些什麼？1899 年，在一味稱作“龍骨”的殘片上，他看出上面刻有一些符號。從事金石文字研究多年，王懿榮知道其中必有名堂。派人到藥店打聽後，他得知“龍骨”來自河南。於仔細端詳店裡的龍骨碎片後，他驚喜地發現，上面刻的符號，很像遠古的象形文字。他能辨識出若干，且看出有些是商朝帝王的名字。他判斷那批龍骨是商朝占卜用的，古人將文字刻寫於龜甲或獸骨上，巫師觀察透過燒灼後的裂紋，來占卜凶吉。王懿榮發現藥店裡龍骨上的符號，是比“鐘鼎文”更古老的文字。由於是刻在龜甲或獸骨，因而此種古文字，後來被稱為“甲骨文”。

如同菜單上的虎掌那道菜，食材中並非真的有老虎之手掌，何況也沒那麼多老虎可供食用，而是取豬蹄筋或膝韌帶部位，龍骨也非真的是古代龍之骸骨的化石，而就是大型哺乳動物的化石。但怎會其中也有龜甲？古時大型獸骨及龜甲，均適合在上刻字，以做為占卜用。而龜與龍都屬四靈(即麟、鳳、龍、龜)，龜又極長壽，所謂千年烏龜，於是不但幾千年後，由地底挖掘出來的獸骨化石，連同混在其中的龜甲化石，都被藥商拿去當藥材，一律名之為“龍骨”。至於“鐘鼎文”又是什麼？鐘鼎文亦稱為金文，乃古代鑄刻於青銅器上的文字，始創於商朝中期，但盛行於西周。主要記載當時王公貴族的重要活動，因其鑄刻於金屬容器上，故稱為金文(亦稱銘文)。商、周二朝盛行青銅器，而青銅禮器以“鼎”為代表，樂器則以“鐘”為代表，於是金文又名“鐘鼎文”。

自發現“龍骨”裡有玄機後，王懿榮即透過各種管道，

廣泛收集刻有字的“龍骨”。只是發現後的隔年(1990 年)，王懿榮在八國聯軍攻入北京時殉國，因此他並未能對甲骨上的古文字，有太多探討，之後他所收藏的那 1 千多片甲骨，落到劉鶚(1857−1909)手上。劉鶚字鐵雲，他即著名的“老殘遊記”一書之作者。他收藏甲骨始於 1901 年，共擁有 5 千餘片。他將所有有字甲骨，一片一片拓下來，並從中精選出 1,058 片。1903 年，史上第一部(共 6 冊)有關甲骨文的著作“鐵雲藏龜”刊印出版。在書中，劉鶚說甲骨文是“殷人刀筆文字”，這是第一份明確指出甲骨文的使用時代為商朝之文獻。王懿榮應也有此認知，但口說無憑，學術上對於創見提出的先後，一般乃依著作發表的早晚。劉鶚對甲骨文，其實也沒有深入的研究，且他書上對甲骨文的分類，亦不是很有系統。不過如果說王懿榮是首位理解到甲骨文之學術價值者，則引發更多人參與甲骨文研究的興趣，便應是“鐵雲藏龜”之出版了。1909 年，劉鶚病逝於新疆後，他所收藏的甲骨，也就流落四方了。

1904 年，語言學家孫詒讓(1848−1908)，依據“鐵雲藏龜”完成“契文舉例”。這是第一部考釋甲骨文的著作，全書約考釋 334 個字，後來被判定正確的約有 185 個。書稿完成後，孫詒讓請人抄寫幾份複本，分別送給劉鶚等人。但直至 1917 年，他過世都已 9 年了，此書才付梓出版。因而孫詒讓對甲骨文的研究，雖屬開山鼻祖，在他生前卻未受到該有的重視。要知學術上，不能只是敝帚自珍，而是得盡力讓著作及時公開發表，以增加影響力。與後人相比，上述 3 位

學者，在甲骨學的研究，成果並不多，但仍是甲骨學篳路藍縷的開創者。1910 年，羅振玉(1866–1940)在發表的“殷商貞卜文字考”中，首度判定甲骨之出土地，乃在“殷墟”。他對照司馬遷(西元前 145–西元前 1 世紀初)“史記”中的“殷本紀”，肯定甲骨乃商朝晚期王室占卜記事用。之後經由考古挖掘，及其他途徑出土的甲骨，估計共約有 10 萬片。從這些甲骨中，已確認有 5 千多個相異字，其中約能辨識 1 千 7 百多個。西周以來，普遍使用青銅金文。據統計，鐘鼎文共約有 3 千字，約可辨識 1 千 8 百多個，比甲骨文略多些。

王國維(1877–1927)，這位“人間詞話”(1910)的作者，有“新學術的開拓者”之稱。大學者陳寅恪(1890–1969)認為其學術成就“幾若無涯岸之可望、轍跡之可尋”。王國維曾以他對甲骨文的研究，來佐證“殷本紀”中，對於商朝的記載，並更正若干錯誤。從以上有關甲骨文之辨識，印證之前所說，考古不僅是挖掘出來便大功告成。一旦挖掘後，便將進行資料分析，並給出推論。但所給的推論是否就為真相呢？並不必然。也許日後新的考古，會給出新的推論，就如前述漢文帝陵寢之新發現。想探索真相，所給推論卻不見得便是真相，這點考古與統計是一樣的。在統計裡，一切都是假設，就看接受那一個。

39 資料探勘(二)

再給一考古裡資料探勘之例，也是有關辨識文字。古埃及的歷史悠久，約在西元前 4000 年，就已出現多個城邦國家(city state)；目前出土的文物中，有約西元前 3400 年前之象形文字(亦稱聖書體(Hieroglyph)，為古埃及的正式書寫體系，其手寫體則稱為僧侶體(hieratic))石板。而埃及的第一王朝，約始於西元前 3150 年，也就是埃及的信史已超過 5 千年。至於舉世聞名的“胡夫金字塔”(Pyramid of Khufu，位於吉薩(Giza)，在開羅(Cairo)西南約 20 里)，此當今唯一倖存的(古)“世界七大奇蹟”(Seven Wonders of the World)，約在西元前 2584 至 2561 年間興建。這座以今日眼光來看，仍屬鬼斧神工的建築，古埃及人可是在距今約 4600 年前完成。由於實在超乎人的理解，至今仍不時有人一口咬定“金字塔是外星人所建”。不過從西元前 11 至西元 7 世紀，約 1 千 6 百年間，埃及先後被幾個帝國征服。在西元前 4 世紀下半葉起，希臘與羅馬人的統治時期，古埃及的文明逐漸沒落。到了西元 641 年，埃及更被新興的阿拉伯勢力完全征服，之後便開始阿拉伯化，人民也被迫改信伊斯蘭教(Islam)。至 12 世紀，埃及已普遍使用阿拉伯語文(至今仍為埃及的官方語文)，延綿數千年的古埃及文明，遂被阿拉伯文明所取代。而埃及人對其祖先輝煌的歷史，也就幾乎一無所知了。比較一下，中國的信史，始自約 3500 年前的商朝，遠沒有埃及早。不過即使“中國文明是世界唯一未曾中斷的文明”，此說法

雖不盡然正確，但 3500 年來，中國就算曾數度被外族入侵甚至征服，但中原民族的主體語文，並未根本改變或消失，且以傳統漢語文寫成的古典文獻，保存至今。

古希臘馬其頓的亞歷山大大帝((Alexander the Great，西元前 356−323 年)過世後，他建立的橫跨歐亞非之大帝國迅速瓦解，被他的幾個部屬瓜分。埃及歸托勒密王朝(Ptolemaic dynasty，西元前 305−西元 30 年)，由托勒密一世(Ptolemy I Soter，西元前 367−282 年，西元前 305−282 年在位)所開創。托勒密王朝的歷任君主，皆為希臘人的後裔，因而一直習於使用希臘語文。雖統治埃及，但王室幾乎都對學習埃及語文毫無興趣。不過官方文件，仍採希臘文與埃及文並存。到了 1 世紀，古埃及的象形文字便不再通行了。基督教的興起，一神論的主張，更迫使埃及人不再興建被羅馬人視為“異教”的神廟。原本建造神廟時，會在牆壁及梁柱等處，刻上象形文字，敘述一些重要事蹟。如今這需求沒有了，於是象形文字漸趨式微，而已存在神廟裡幾千年，無數的象形文字雕刻，遂逐漸形同廢文，只能當成藝術品看待了。現存最後一個標示有日期的神廟上之象形刻文，時間是西元 398 年 8 日 24 日。此後便不再有了，因象形文字，刻或寫給誰看呢？約到 5 世紀後，古埃及所留下大量又重要的文件，便再也無人看得懂了。誰能讓那些從 5 世紀起，便被視為藝術品的象形文字，成為有意義的文字呢？

法國學者商博良(尚−法蘭索瓦商博良，Jean-Francois Champollion，1790−1832)，在語言方面極具天賦，精通多種

文字，且對科普特(Coptic)文，特別感興趣，那是西元 1 世紀後，埃及所創造出之文字。在 18 至 19 世紀初，當時考古學僅在萌芽階段，尚非一能嚴肅看待的科學。“聖經”的“創世紀”裡，有描述一些人類早期的歷史，只是那畢竟是以色列人所記，人類的誕生，真是那樣嗎？古埃及那些看起來神秘無比的象形文字中，是否有關於世界起源的記載，可用來與“聖經”相互印證？令人好奇。這是何以那時對人類世界起源有興趣者，讀了“聖經”後，不少會對古埃及的文化，產生濃厚興趣的原因之一。但由於象形文字無人能識，以致對古埃及的了解，當時幾乎一片空白。1809 年，商博良 19 歲時，決定投入科普特文的學習，他想藉由科普特文，來通向古埃及文。

1798 年 5 月，拿破崙(Napoleon Bonaparte，1769−1821)率軍出征埃及，與遠征軍隨行之考察團，包含約 175 位各領域的學者，預計進行各項研究。埃及人習於立碑，凡有什麼大事，便立個碑刻上事件經過。古物的挖掘，亦是考察團此行的目的之一。法軍在羅塞塔(Rosetta，位於亞歷山大城以東約 65 公里)，發現一塊石碑，其上刻有 3 段文字，乃 3 種可對照的字體：希臘文、古埃及象形文字，及古埃及通俗文(亦稱世俗體(Demotic)，由僧侶體演變而成)。此 3 種文字，皆為托勒密王朝時流通的文字。經翻譯希臘文後，得知這是西元前 196 年 3 月 27 日，孟斐斯(Memphis，位於開羅南方 20 公里處，曾是古埃及的首都)祭司所撰寫的敕令。記述托勒密五世(Ptolemy V Soter，西元前 209−181 年，西元前 204−181 年

在位)的即位慶典。

此塊在拿破崙時代，已有將近兩千年歷史的石碑，隨即被運至開羅。法國科學家興奮莫名，當下進行文字拓印、素描，及鑄模，且做出多種複製品，並開始研究。他們樂觀地以為，象形文字的解讀，已指日可待。只是並沒那麼容易，還要再經過至少 20 年的艱辛鑽研，才能破解。1801 年，法軍敗於英軍。英國本要將法國所有的研究成果，以及那塊令他們垂涎不已的石碑都沒收。寧為玉碎不為瓦全，法國學者威脅要毀去一切，包括石碑。最後英國不得不讓步，同意法國可保留研究資料及石碑拓片等，他們只奪走石碑。法國人並沒什麼好抱怨的，因向來弱肉強食，何況一切都沒問埃及人的意見。後來石碑便以發現地，命名為羅塞塔石碑(Rosetta Stone)，今日在大英博物館展示。

古埃及人以莎草紙(papyrus，以紙莎草(apyrus sedge)的莖製成)來書寫，這種文書便稱為“紙草文書”。古埃及人遺留下大量的“紙草文書”，沙漠地區乾燥的氣候，讓文書能長期保存良好，但至 19 世紀初，已長達 1 千多年無人能辨識。商博良從 1822 至 1824 年間，全心全力投入羅塞塔石碑上象形文字的研究。他熟悉希臘文，持續相互比對羅塞塔石碑上的 3 種字體，並運用他的科普特文知識，終於逐漸能了解某些古埃及通俗文字之意義。此外，他閱讀無數“紙草文書”，加上不斷推測，有如在解密碼。1822 年，他成功地譯解出古埃及象形文字的結構，並編出完整的埃及文字符號和希臘字母的對照表。長達 3 千年，無數古埃及的歷史文獻，

終於可一一解開了。

1824 年 6 月，商博良經阿爾卑斯山(Alps)，前往義大利北部杜林(Turin)，那裡有大量古埃及象形文字文獻的收藏。他找到一份莎草紙文書，那是拉美西斯二世(Ramesses II，約西元前 1303−1213 年，約西元前 1279−1213 年在位)時製的。由於其上列有在拉美西斯二世之前，歷代法老的名字，後來便以發現地，將此份文書命名為“杜林國王名冊”(Turin King List)。雖有若干脫漏，但在拉美西斯二世以前，3 百多位埃及統治者的名字，及各自在位的時期，仍大致知道了。此對了解古埃及史，助益極大。對商博良的研究過程有興趣者，不妨去看“破解古埃及”(The Keys of Egypt: The Race to Read Hieroglyphs，2001，萊斯利亞京斯與羅伊亞京斯(Lesley Adkins and Roy Adkins)著，黃中憲(2007)為中譯本)一書。

不要說探勘不見得必有所獲，人們常說大數據，但有了大數據，一切就能迎刃而解嗎？對“殷墟”出土的那約 10 萬片的甲骨，雖文字的演變多少有些脈絡可尋，但已確認的 5 千多個相異字中，能辨識的也不過 1 千 7 百多字，才佔約 1/3。其他約 2/3 就束手無策了。古埃及人留下不可計數之“紙草文書”，大數據在那裡。但經反覆比對，就能撥開謎網，讀懂古埃及文了嗎？事實上，19 世紀初，英法等國同時有多人致力於破解工作，只是進展一直很有限。直到羅塞塔石碑出土，而將閱讀古埃及文，視為一生懸命的語言天才商博良，看出如何利用此石碑，辨識古埃及文，終於得以破解。所以，並非任誰有了大數據，便有了解密鑰匙，無往不利。

其他像人工智慧(Artificial intelligence，縮寫 AI)也是一樣。眾所皆知，獲年度總冠軍，為 NBA 各隊最高目標，但付出天價年薪，聚集幾個頂尖球員，得總冠軍便能如探囊取物嗎？顯然不一定。同樣地，也非利用電腦，便能有支智慧棒，指揮一切。自己須先有智慧才行。

因覺身上某部位不太對勁，到醫院看診後，醫生指示做影像檢查，諸如照 X 光、超音波、電腦斷層掃描(Computed Tomography，縮寫 CT)、核磁共振造影(Magnetic Resonance Imaging，縮寫 MRI)、正子斷層造影(Positron Emission Tomography，縮寫 PET)等。得到一些影像後，要有專人判讀，了解病因，以對症治療。影像判讀會有失誤嗎？在資料眾多下，或許難免吧！利用統計，藉由分析歷來大量影像資料之判讀結果，可對每一就診者之影像，給出建議。此即醫療人工智慧，能當做醫師判讀影像之輔助，目前已有多家醫院在進行此類研究計畫。