談統計素養

黃文璋

1 統計難以數學化

教育部在“十二年國民基本教育實施計畫”中，提出五個國民素養，包括語文(含中文及英文)、數學、科學、數位，及教養/美感。此為 18 歲的國民，於高中畢業時，所需具備的五大基本能力。之後在“提昇國民素養實施方案”的“數學素養研究計畫”之結案報告裡，訂出四大數學知識素養領域：變化與關係、空間與形狀、數量，及不確定性與數據。數學成為現代國民須具備的五個素養之一，顯示國民素養裡，對數學的重視。又由於不確定性與數據，為統計學的範疇，顯示數學素養裡，對統計的重視。

一直以來，在分類時，統計學常是放在數學中，而不像物理、化學，及生物，是在一更大的領域，即科學之下。在生物分類法(taxonomy)裡，從大至小，有界、門、綱、目、科、屬、種。看起來完全不一樣的獅子、老虎，及豹等動物，都屬於貓科。但這些動物，在有關動物的書裡，乃分別介紹，而非將獅子放在貓裡介紹，獅子與貓的特性畢竟有相當大的差異。如果只是如生物的分類，則將統計學與數學歸於同一類，那就無妨。但如今是在數學課程中學統計，那便是視統計學為數學。要知物理雖用到不少數學，但物理並無法放在數學課程中教。統計學也一樣，是用到許多數學，但統計與

數學所注重的，乃大不相同。若只以數學的眼光來看統計，大約只會看到其中的數學部分，很難看清楚全貌。因此多年以來，中學裡的統計教學，總是顯得左支右絀，相當彆扭。雖持續修訂課綱，將不適合的題材相繼取消，且提供許多補充說明，卻仍爭議不斷。追根究柢，乃統計難以數學化。只要是放在數學裡的統計題材，便不容易被正確對待，更遑論讓不確定性與數據，成為國民的基本素養。

2 連百分位數都搞不定

原本在“九年一貫數學課程綱要”(底下簡稱“九年課綱”)裡，百分位數(percentile)的相關題材，乃列在統計與機率的主題中，被置於九年級(即國中三年級)。近年於制訂“十二年國教數學課程綱要”(底下簡稱“十二年課綱”)時，部分百分位數的內容，被移出國中數學了。更明確地說，在“十二年課綱”裡，在七年級(即國中一年級)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 Σ 鍵計算平均數；

在九年級有：

> 統計數據的分布：全距；四分位距；盒狀圖。

國中數學已不見百分位數了。但仍有中位數，那是一種特別的百分位數。又在普通高中一年級(即十年級)，於數據分析的主題下有：

> 數據分析：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。

其中並未提到百分位數，百分位數自此從中小學數學中消失了嗎？

百分位數是什麼？既然是“百”，你腦海中大約會立即浮現出一組數據 1，2，…，100。直觀上，1 就是第 1 百分位數，2 就是第 2 百分位數，…，100 就是第 100 百分位數。那中位數呢？數據中間有兩個數 50 及 51，要嘛都當做中位數，要嘛取二者之平均，即 50.5 亦可，怎麼規定就怎麼做。百分位數看起來一點都不難，放在國中好好的，何以只留下中位數，至於一般的百分位數便不教了？

百分位數其實移到高中了。你可能感到納悶，前面所引“十二年課綱”裡，那段“數據分析：…標準化”，其中並未出現百分位數的字眼。況且，高中生學這種東西，豈不太簡單？

百分位數是一個大家耳熟能詳的名詞，即使沒學過也大致知道意思。但放在數學課程中，便須有明確的定義。先看百分位數。“九年課綱”裡並未給百分位數的定義，在“九年級分年細目”中，只寫著：

> 能以中位數、四分位數、百分位數，來認識資料在群體中的相對位置。

沒有更多，就僅這麼一句話。事實上，對每個概念，課綱裡通常都寫得很簡短，連微言大義都談不上。之後在“附錄一”中的“九年級細目銓釋”(底下簡稱“銓釋”)中，則有如下說明：

> 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。
>
> 百分位數和中位數、四分位數一樣，可以表示某資料組在總資料中的相對位置。學生應能自資料之相對累積次數分配表求出百分位數。
>
> 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。

又在“附錄四”“標準用詞與解釋”(底下簡稱“解釋”)，於“中位數”項下是：

> 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。

於“百分位數”項下則是：

> 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。

如前，假設有數據 1，2，…，100。則依“解釋”，第 0 百分位數為 1，第 1 百分位數為 2，…，第 99 百分位數為

100，跟人們之前想的並不一樣。至於中位數，由於是第 50 百分位數，故為 51。51 是中位數，它的左鄰 50 卻非中位數，頗令人不解。還有，比 51 小的數有 50 個，的確佔一半；但比 51 大的數有 49 個，僅佔 49%，並未達一半。怎會這樣？不只如此，檢視“銓釋”，51 不但不在前後各切一半的中間位置，前後兩邊的個數也不平衡。尚有一點令人疑惑，即能否有第 0 百分位數？我們不知道，在“九年課綱”裡，從頭到尾都沒說。但若沒有，則數據中的第 1 個，也就是 1，便不是任何百分位數，那將很奇怪。所以我們當做有。

不妨換組數據來看，假設有數據 1，2，…，99。則第 0 百分位數仍為 1。第 1 百分位數呢？不是 2 嗎？比 2 小的數只有 1 個，佔 1/99，並非 1/100，所以 2 不是第 1 百分位數。那 2 是第多少百分位數？依“標準用詞與解釋”，就是第 1/99，或說第 1.0101…百分位數。只是能否有非整數的百分位數？“九年課綱”裡雖沒說，但理論上可以有。事實上，不但沒有第 1 百分位數，連第 2，…，第 99 百分位數都不存在。而既然第 50 百分位數不存在，也就不存在中位數。但長久以來，對這組個數為奇數的數據，人們不是認為正中間那個 50，就是中位數嗎？雖然比 50 小的數有 49 個，比 50 大的數也有 49 個，均並未各佔一半。

上述兩個例子顯示，有關百分位數的“銓釋”及“解釋”，充滿矛盾。一直以來，在教學及學習上，此單元想必帶給師生相當多的困擾。這樣的學百分位數，如何能讓學生“認識資料在群體中的相對位置”？更不可能體會到什麼

“數學之美”了。

等等!有人警覺到了，會不會是我們所舉之例，皆屬於“資料數少”，該避免拿來當例子？當然不是。諸位不妨自行舉例，很快就可看出，對於前述那些問題，即使資料數再多，也仍存在。你現在應知道了，在制訂“十二年課綱”時，何以負責國中數學的委員，要建議將百分位數這塊燙手山芋移出。令人好奇的是，難道當初寫那些“銓釋”及“解釋”的“九年課綱”委員，都是閉門造車，而沒順手給幾個例子，看這樣的銓釋或解釋，是否會窒礙難行？這我們就不知道了。說不定是覺得太簡單，所以沒有多加留意。

前述提到“十二年課綱”裡，雖未述及“百分位數”，但其實內含百分位數。因在“說明手冊”中，提到“新課綱將過去九年級的百分位數移到此”。原來課綱中沒列出的，並不表示就沒有涵蓋，真有夠隱晦的。只是不免擔心，百分位數放進高中數學，之前我們所指出的那些問題，就能迎刃而解嗎？有人可能還好奇，中位數為一種特別的百分位數，何以中位數能放進國中，百分位數卻不行？

對百分位數，有人提出如下的“補充說明”：

> 第 m 百分位數 P_m，指的是同時滿足：小於等於 P_m 的資料至少占全部資料的 m%以上，大於等於 P_m 的資料至少占全部資料的(100-m)%以上。

有些人則除上述條件外，建議增加一點補充：

> 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。

我們來檢視上述“補充說明”。首先，“九年課綱”中使用的“占”，被改成“佔”了。為了本文前後一致，我們仍採“佔”。另外，“補充說明”中的“小於等於”，宜寫成“小於或等於”。仍先以數據 1，2，…，100 為例。則依“補充說明”的第一點，得 P_0=1，P_1=1，2，P_2=2，3，…，P_{50}=50，51，…，P_{99}=99，100，P_{100}=100。這樣的規定，使中位數 P_{50} 有 50 及 51 兩個值，尚可接受。但除 P_0 及 P_{100} 外，每一 P_m 皆有兩個值，m=1，2，…，99。而每一 m=1，2，…，100，也皆等於兩個百分位數，即 $m=P_{m-1}=P_m$。第 1 百分位數不只是 1，也可以是 2，第 2 百分位數則除 2 之外，也可為 3，…，此明顯違反一般人的認知，疑惑自然產生，這是何以會有人建議增加一點“補充說明”的原因。但加上此點補充後，P_0=1，P_1=1.5，P_2=2.5，…，P_{50}=50.5，…，P_{99}=99.5，P_{100}=100。中位數=50.5 仍沒問題，但第 1 百分位數不是 1，而是 1.5；第 2 百分位數不是 2，而是 2.5，…，與一般人所想的完全不同。這種百分位數，能被接受嗎？

其次看數據 1，2，…，50。則依“補充說明”的第一點，得 P_0=1，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，P_{100}=50。中位數 P_{50} 有 25 及 26 兩個值，可以接受。當 m 是 0、100 及奇數時，P_m 都只有一個值；當 m 是偶數時，P_m 皆有兩個值，m=2，…，98。若

再遵循第二點“補充說明”，則得 $P_0=1$，$P_1=1$，$P_2=1.5$，$P_3=2$，$P_4=2.5$，⋯，$P_{50}=25$，⋯，$P_{98}=49.5$，$P_{99}=50$，$P_{100}=50$。中位數仍沒問題，但其他百分位數，恐讓學生看得糊里糊塗。

問題並不只上述那些，我們再看一例。假設美國某種球的職業球員共有 1 千位。按年薪由高至低排序，前 9 位年薪各 3 千萬(美元)，第 10 位及 11 位年薪各 2 千萬(美元)。各位不妨自行依“補充說明”，立即可得 $P_1=2$ 千萬。3 千萬呢? 依“補充說明”，那 9 個 3 千萬，對 P_1 毫無影響。至於沒學過“補充說明”的人，只能憑直觀行事。1 千的 1%為 10，前 10 位最高薪球員的平均年薪為 2.9 千萬，即得 $P_1=2.9$ 千萬。不同的定義導致不同的 P_1，究竟那一較合理? 就留給讀者自行探索。

在“十二年課綱”中，對百分位數所給的“補充說明”，不像之前九年課綱中所給的“銓釋”及“解釋”，數學上並沒有瑕疵。對任一組數據，每一百分位數皆可明確求出。但連 1，2，⋯，100，這麼四平八穩的數據，依一點或兩點“補充說明”，得到的每一百分位數，不是兩個，就是 1.5，2.5 之類的，將讓學生摸不著頭緒。升學至上，多想無益，最後恐怕只好把百分位數當做(簡單的)數學來學。豈會想到與數據分析有何相干？豈會明白百分位數有有什麼大用？

中位數應是百分位數裡，人們較常接觸到的。平均成績、平均國民所得等，若想以一個單一的值，來代表一組數據，平均數常被採用。但也有一些情況，平均數並不那麼適

合用來當代表值。在職業球隊，球員薪資差異往往很大。大部分的球員，薪資都不太高，少數較高，且每隊總有一兩位薪資是天價的球員，會將全隊球員的平均薪資，大幅拉高。這時光看平均薪資，可能使人們誤以為球員薪資普遍很高。由於少數極端值，並不影響中位數之值，因此當數據中有較極端的值時，中位數常較平均數更適合用來當代表值。中位數大致是位於一組數據中間的值，前後約有各半的數據。媒體偶有關於職業球隊薪資的報導，可能因涉及球隊財務，及球員隱私等原因，數據常無法太精準。例如，曾有新聞說美國職業棒球大聯盟(Major League Baseball，縮寫 MLB)，2015 年球員年薪的中位數是 470 萬美元。怎麼那麼粗糙，只計到 10 萬？因球員很年從球隊實際得到的總薪資，有時包含績效獎金等，說來有點複雜。而只不過想讓人們對球員年薪的多寡，能約略有些概念。這時太在乎細節，就不是那麼必要。100 萬跟 470 萬是有差，至於 470 萬跟 478.5 萬，又不是自己的薪資，有幾個人會很介意？因此宣稱的中位數，是真的正中間那個值嗎？或者是某幾個的平均？何須太計較？只要是一差不多位在中間的值，就可以了。

百分位數的情況類似。假設政府公佈 2016 年，台灣國民所得的第 5 百分位數(即 P_5)。怎樣叫有所得？學生打工，及擺地攤者，也都算嗎？而凡有所得的人，及其全部所得之數，政府真能精確掌握？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，相信差異很大。因此其中百分位數到底如何產生，何須太在意?反正就是讓人大概知道，全國最高收入的前 5%，究竟高到那裡去。就算數據

有些含混，也不必挑剔，或企圖追根究柢，因僅供參考而已。

在很多實務中，百分位數不過就是差不多之事，不必深究，事實上也無法深究。這樣不求弄太清楚的題材，豈適合出現於在乎定義、講求準確之中小學數學中？

3 百分位數的應用

“九年課綱” 中強調：

> 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。

大部分的人，對這點應可理解。當資料數過少，百分位數便的確不太能發揮作用。例如，假設有數據 1，2，…，10。則依“十二年課綱”，可得 $P_1=\cdots=P_9=1$，$P_{10}=1$，2；$P_{11}=P_{12}=\cdots=P_{19}=2$，$P_{20}=2$，3；…；$P_{91}=P_{92}=\cdots=P_{100}=10$。若再依“補充說明”的第二點，可得 $P_{10}=1.5$，…，$P_{90}=9.5$。1 至 10，每個數同時為 9 或 10 個百分位數。這時百分位數，對了解每一資料，在全體中的相對位置，並無太大幫助。再看一例。假設有一筆數據，先是 30,000 個 1，接著是 30,000 個 2，然後 3，4，…，10，各有 5,000 個，總共有 100,000 筆資料，資料數不少。結果 $P_1=\cdots=P_{29}=1$，而 $P_{30}=1$，2，或者取 $P_{30}=1.5$；其他百分位數亦不難得到，就讓各位自行完成。相信大部分的人也一目瞭然，看似資料數有 10 萬之多，因重覆之故，實際相異的資料數只有 10。於是 1 至 10，每

個數仍同時為多個百分位數。“大學入學考試中心”(簡稱“大考中心”),每年負責統籌舉辦的“大學學科能力測驗”(簡稱學測),各科考生,都有好幾萬人,因此百分位數,可在其中運用自如?

之前我們已指出,百分位數這種題材,並不適合出現在中小學數學裡。只是與高中生關係密切的學測裡,也用到百分位數,有人可能因此會說,所以中學生還是該懂得百分位數。百分位數為何會出現在學測裡?

台灣有學測,始自民國 83 年開始實施“大學多元入學新方案”,至今已有不短的歷史。其中細節,二十餘年來,曾經數度修正。高中生想進大學,學測成績相當關鍵。學測共有國文、英文、數學、社會及自然等 5 科。每科原始滿分依序是 108、100、100、144,及 128。其中有 3 科的原始滿分,都不是常見的 100,似乎是各科專家深思熟慮後所產生。但沒有用,因後來“大考中心”會將考生的各科成績,皆轉換成 0 至 15 的“級分”(83 學年度只有 10 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”。這理由很可笑,原始成績若只少 0.1 分,並沒什麼大差別。如今卻可能差了很顯著的 1 級分。原始總分較高者,總級分卻可能少 4 級分。

級分如何得出?各科取前 1%考生成績的平均,不妨以 a 表之,將 a 除以 15,$a/15$ 即為各級分之級距。原始成績至少有 a 的,為 15 級分;低於 a,但至少有 $14a/15$ 的,為 14 級分,餘類推,至於成績低於 $a/15$ 的,則為 0 級分。由於科

目之別，再加上不同年度題目難易之差，成績散布之變異可能很大。導致科目或年度不同，級分無法相比。所以不能問 11 級分夠不夠高，或 6 級分算低嗎？如果各科級分是由高至低，依到考人數等分，則各科即使不同年度，同一級分所反映考生的表現優劣，便較接近。

自 91 學年度起，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，及各科級分，依 5 標訂出檢定標準，做為篩選門檻。在此頂標乃成績位於第 88 百分位數之考生級分，前標乃成績位於第 75 百分位數之考生級分，均標乃成績位於第 50 百分位數之考生級分，後標乃成績位於第 25 百分位數之考生級分，底標乃成績位於第 12 百分位數之考生級分。級分是按分數切割，5 標則是按人數切割。忽而分數忽而人數，這套制度設計的邏輯，頗令人難以理解。各科 5 標雖依相同的百分位數，但由於已先經級分的扭曲原始成績，不同科目的同一標，並無法相比。以 106 學年為例，11 級分在國文只是均標，在數學卻是頂標；而數學的均標 6 級分，尚比國文的底標 7 級分還低。

前面看到資料數少時，一數可能同時為好幾個百分位數，會弄得不清不楚。學測考生那麼多，應不會有這個問題吧！這樣想就錯了。考生雖多，可惜成績轉換成 15 級分，連同 0 級分，相當於一筆資料數只有 16 的數據。就像我們

之前所舉，那一資料數有 10 萬之例，每 1 級分，將可能同時為好幾個百分位數。換句話說，學測忽略百分位數該避免用於資料數少的原則，貿然使用。遂造成頂不見得是頂，前不見得是前，均也不見得是均。

學測的 5 標如何求出？“大考中心”在學測的簡章中有說明。假設某科到考生為 161,567 人，分別乘上 5 標對應的百分比，然後取整數，小數部分無條件進位。即得

161,567×0.88=142,178.96→142,179，
161,567×0.75=121,175.25→121,176，
161,567×0.50=80783.5→80,784，
161,567×0.25=40391.75→40,392，
161,567×0.12=19388.04→19,389。

將考生成績，由低至高排序後，從最低分往上數之第 142,179 位到考生的級分，便為頂標，餘類推。此與依“十二年課綱”中的定義，所求出的百分位數，有何不同？如果是數據 1，2，…，161,567，則滿足小於或等於 P_{88} 的資料，至少佔 88%，大於或等於 P_{88} 的資料，至少佔 12%，唯一只有 142,179。同理 121,176 為唯一的 P_{75}，餘類推。即在二定義下，所得的 5 標均相同。既然如此，“十二年課綱”中的百分位數為，何不採學測上的定義就好？事實上，由兩種定義所得之百分位數，並不永遠一致。舉個例子來看。對於數據 1，2，…，100，依學測的定義，得第 10 百分位數為 10，…，第 90 百分位數為 90；但依“十二年課綱”上的定義，得 P_{10}=10，11，或取 P_{10}=10.5；…；P_{90}=90，91，或取 P_{90}=90.5。

各科的級分，是按原始成績等分(以 15 級分為頂)。而基於考題的難易程度，全部考生的原始成績，有不同的集中情況。因此即使同一科，各級分考生數所佔的百分比，可能相差很大。本來每一級分的考生數，平均約有 6%多。但以 104 學年度的國文科為例，12 級分的考生數最多，約占 19.04%；而由 10 至 13，這 4 級分的考生數，便約佔 63.46%；至於 1 級分的考生數才約佔 0.02%，0 級分的考生數才約佔 0.003%。國文科考生集中在某幾個級分的情況，並非僅發生在 104 學年度。不妨來看 105 學年度。考生數最多的是 11 級分，約佔 16.62%，由 9 至 12，這 4 級分的考生數，共約佔 58.77%，而 1 與 0 級分的考生數，仍差不多佔 0%。個人申請入學，於第一階段篩選時，各大學校系可對學測的總級分及各科級分，擇其中若干，依 5 標訂出檢定標準，做為篩選門檻。以國立臺灣大學數學系為例，該系訂出的門檻為，國文均標、英文均標、數學均標、社會後標，及自然均標。“大考中心”規定，於訂門檻時，所能依據的，並非考生的原始成績或級分，而是總級分與各科級分的 5 標。5 標既然在篩選過程中，扮演重要的角色，因此“大考中心”提供之 5 標，所代表的意義，應該要很明確才行。只是並非如此。

如前所述，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。看到這裡，不少人腦海中可能浮現 12、25、50、75，及 88 等百分比，分別為達到各標的考生數之百分比。可惜“大考中心”並無此邏輯。仍以 104 學年度的國文科為例。表 1 給出級分人數百分比累計，其中顯示到考人數為

144,250。

$$144{,}250\times 0.88=126{,}940,$$
$$144{,}250\times 0.75=108{,}187.5\rightarrow 108{,}188,$$
$$144{,}250\times 0.50=72{,}125,$$
$$144{,}250\times 0.25=36{,}062.5\rightarrow 36{,}063,$$
$$144{,}250\times 0.12=17{,}310。$$

從最低分往上數之第 126,940、108,188、72,125、36,063,及 17,310 位到考生的級分,分別為頂標、前標、均標、後標,及底標。由表 1,即得分別 13、13、11、10,及 8 級分。頂標與前標,居然同為 13 級分!某大學中文系,每年在訂申請入學各科的篩選門檻時,斟酌再三。少子化導致報考人數年年下降,先得預測今年申請本系的人數,然後決定國文科究竟該訂頂標或前標,以對本系較有利。結果看到“大考中心”公布的成績統計後,發現花大功夫的討論全白費了。國文科的頂標及前標,二標竟然沒有差別。怎會這樣?

我們已指出,各科的原始成績只是過客,皆被“大考中心”轉換成級分。由於各科的級分,總共才 16 筆相異資料,而對一筆資料數不多的數據,一數可能同時為好幾個百分位數。“大考中心”看似經過精心規劃,所產生的百分比:88、75、50、25,及 12,最後完全不是那麼一回事。這說來並不奇怪。先看頂標,由表 1,國文科成績位於第 88 百分位數,即由最低分往上數起,第 120,964 位到考生,其級分為 13。但與該考生,同為第 88 百分位數的考生,共有 23,824 位。而 13 級分以上(即 13、14、15 級分)的考生,共約佔 28.50%。

達到頂標的考生，不是該僅佔 12%嗎?如今卻佔了超過 4 分之 1！因此國文科拿 13 級分，只能算是考在前面，怎麼會是什麼“頂級的標準”？

再從另一方式來看。由表 1，國文科 14 及 15 級分，共約佔 11.98%，未達 12%。雖才差約 0.02%，但按其算法，要往下降 1 級分。但 13 級分的考生有 16.52%。以 0.02%換 16.52%，這樣一來，達到頂標的考生，不但遠超過 12%，甚至超過前標的下界 25%，有 28.50%。於是頂標與前標，便同一級分了！我們來看，由表 1，成績位於第 75 百分位數，即由最低分往上數起，第 108,188 位到考生，其級分正是 13。表 2 給出，104 學年度學測各科及總級分，5 標的級分，及達到的考生，所佔之百分比。達到及設定的百分比，有些的確差異很大。仍看國文科。不只頂標考生佔的百分比超多，達到均標的考生，有 63.81%，比設定的 50%，高出 13.81%之多。這都是因級分數太少之故，或者說將原始成績轉換成級分之故。

不同標卻同一級分，除 104 學年度國文科外，尚有 96 學年度的國文科，及 92 學年度的社會科。3 次皆是頂標與前標同為 13 級分。當資料數少時，百分位數在教學上，該避免拿來做為例子。這點即使課綱沒提醒，也應是常識。連當例子都不適合，何況在攸關全國高中生進大學的比序，資料數才 16 而已，怎宜引進百分位數？“大考中心”的專家沒學過百分位數嗎？當然不至於。只是如果連他們學了都沒用，中學生學了又何用？

雖然同一級分，對不同科目或不同年度，並無法相比，這是現行級分制一很大的缺失，但即使專業人士，也不見得很清楚這點。2017 年 3 月 16 日，聯合報有一則關於大一新生國文課教學的報導。其中台灣師範大學國文系指出，該校每年新生有 5 成以上，國文學測至少有 13 級分，而 10 級分以上的比率，更高達 9 成，代表不論文、理組，國文程度都有一定的水準。真的是這樣嗎?以 104 學年度為例。10 級分夠不夠好?而 13 級分很高嗎?由表 2，國文 13 級分以上，約佔 28.5%，超過 4 分之 1;而 10 級分以上，則約佔 75.45%，多達 4 分之 3。所以拿到 13 級分，算是中上;而 10 級分，恐怕只能說屬於中下了。至於數學，10 級分以上才佔 25.59%，好過國文的 13 級分;再由“大考中心”的網站，可得 13 級分以上，僅佔 8.34%。所以，同是 104 學年度，數學考到 10 級分以上，尚可宣稱都具有一定的水準;而考到 13 級分以上，沾沾自喜並無妨。但其他科目，便可能是另一回事了。

人們常講大數據，以為一旦有大數據，一切大小問題，均可迎刃而解。如今明明是小數據，卻視為大數據，然後隨興揮舞百分位數。可見若缺乏數據素養，真有大數據，豈有大用？不過暴殄天物而已。

比起眾多深奧的統計方法，百分位數縱有不同的定義，但概念皆屬淺顯，理解不難。實際應用時，要切記這是有關數據分析，而非在做算術練習，只管求值即可。百分位數是統計，統計不可數學化。若只在乎數學，將難培養出統計素養。

有關百分位數進一步的參考資料，可見黃文璋(2014a)、(2015a)，及(2015b)三文。

表 1 104 學年度學測國文科級分人數百分比累計表

級分	國文			
	人數	百分比	累計人數	累計百分比
15	4,402	3.05	144,250	100.00
14	12,882	8.93	139,848	96.95
13	23,824	16.52	126,966	88.02
12	27,459	19.04	103,142	71.50
11	23,486	16.28	75,683	52.47
10	16,778	11.63	52,197	36.19
9	11,015	7.64	35,419	24.55
8	7,548	5.23	24,404	16.92
7	5,475	3.80	16,856	11.69
6	4,185	2.90	11,381	7.89
5	3,165	2.19	7,196	4.99
4	2,223	1.54	4,031	2.79
3	1,356	0.94	1,808	1.25
2	417	0.29	452	0.31
1	30	0.02	35	0.02
0	5	0.00	5	0.00

表 2 104 學年度學測各科及總級分 5 標一覽表

標準 項目	頂標	前標	均標	後標	底標
國文	13(28.50%)	13(28.50%)	11(63.81%)	10(75.45%)	8(88.31%)
英文	14(12.18%)	12(30.14%)	9(55.73%)	6(75.49%)	4(91.94%)
數學	12(12.98%)	10(25.59%)	7(51.27%)	4(82.03%)	3(90.06%)
社會	14(16.63%)	13(30.11%)	11(60.90%)	9(80.59%)	7(91.76%)
自然	13(15.92%)	11(31.38%)	9(50.30%)	6(80.75)	5(90.39%)
總級分	63(13.27%)	57(25.86%)	47(52.08%)	36(76.75%)	28(88.24%)

註. 各級分括號中之百分比，為達到該級分之累積考生所占百分比。

4 取樣

“不確定性與數據” 為國民 4 大數學素養之一。其中 “不確定性” 一詞，我們認為以 “隨機性” 取代，或許較恰當些。因一般說到不確定性，常有點負面的意思。如政治裡向來就有各種大小的風險存在，對不確定性的存在，本不該

感到訝異。但 2017 年 1 月 10 日，經濟日報有一則報導：

> 英國準備脫離歐盟，美國新政府的貿易政策可能衝擊中美關係，政治不確定性引發市場避險需求，帶動黃金價格五個交易日來第四次上漲。…。

顯示不確定性讓人排斥。再給一例。橋本忍(1918−)是日本著名的電影編劇家，長期與大導演黑澤明(1910−1998)合作，兩人共同完成不少好作品。在他(2006a)一書中，談論多部黑澤明拍攝的電影。如同一流球員，再怎麼了不起的導演，也有失手的時候。在此書便提到：

> “生者的紀錄”(1955)的失敗，是因為劇本是帶有未知數與不確定性的原創故事，…。

將拍片失敗，歸罪於劇本裡有未知數與不確定性。在黃文璋(2014b)一文，尚舉了一些不確定性，讓人不喜，避之唯恐不及的例子。

處在此隨機世界裡，隨機性導致不確定，而在不確定性下，自然有很多事屬未知。但事實上，雖說隨機，可不等同於隨意，仍會遵循某些法則，並非完全不可預期。因此，與其擔心不確定性，不如去了解隨機法則，因而設法做出較佳的決策。而就像製磚不能沒有黏土，欲有好的決策，數據常是不能少的。分析數據，從中挖掘出一些可掌握的資訊，以讓不確定性裡，不都那麼不確定。這是何以數學素養裡，不確定性要與數據結合在一起。

高中數學自“九五暫綱”(民國九十五年開始實施的高中數學課綱)，引進“信賴區間與信心水準的解讀”。為什麼高中生要學信賴區間？媒體上，經常有關於民調的報導，其中往往附有信賴區間，信賴區間儼然是國民該具備的統計素養，遂堂而皇之地進入高中數學。對某議題，光給樣本顯示的整體支持率還不夠，亦想了解不同族群的支持率之差別之於是民調裡常會做的“交叉分析”，也就跟著進入高中數學。至於主要為量測二變數間的線性關係之強弱及正負的“相關係數”，也被高中數學招手進入了。可以這麼說，自“九五暫綱”起，以數學家為主的課綱委員，覺得高中生該多懂些統計，遂增加高中數學裡的統計分量。其後“九九課綱”中，取消交叉分析，“十二年課綱”中，則連信賴區間也取消了。十年來，信賴區間給高中師生帶來的困擾，及掀起的爭議，即將煙消雲散。只是信賴區間真有那麼難學？

在金庸(1924–)“鹿鼎記”(1972)一書的第二十二回，對少林寺武功一竅不通的韋小寶，要澄觀教他一指禪。澄觀是少林寺般若堂的首座，武學所知之博，被寺中群僧，推為當世第一。澄觀說：

> 咱們少林派武功循序漸進，入門之後先學少林長拳，熟習之後，再學羅漢拳，然後學伏虎拳，內功外功有相當根柢了，可以學韋陀掌。

預備功夫可真多。還沒完，韋陀掌後，依序是散花掌、波羅蜜手、金剛神掌、拈花擒拿手、般若掌、易筋經，待這些都練熟了後，才能學一指禪。澄觀自稱練了 42 年，才略窺門

徑，在少林寺千餘年來，名列第三。最快的那位，花了 36 年，次快者則花了 39 年。

高中數學裡的信賴區間，乃一綜合性的統計題材，其中包含好幾個統計主題，至少有取樣、估計、誤差，及極限；涉及的分佈有二項、超幾何，及常態；包含的概念有隨機性、機率的意義，及條件機率。信賴區間若能弄懂，表示統計已有相當基礎了。一般在大學統計學的教科書裡，信賴區間通常置於全書的後半部。而交叉分析出現時，全書可能已到尾聲了。又由於要用到近似，需要藉助中央極限定理，或者僅是較簡單的版本，即二項分佈的常態近似。但此機率裡相當重要的極限定理，即使是初步的版本，都沒太多大學數學系的畢業生，能講得清楚明白。雖在大學統計課程裡，於足夠的鋪陳後，才會引出信賴區間。而就算這樣，大學生對信賴區間的涵義，仍不過一知半解而已。如今企圖在高中數學很少的篇幅中，講授信賴區間、中央極限定理及交叉分析，彷彿想在短時間內，教會韋小寶一指禪，師生皆備嘗辛苦，乃是必然。

先看取樣。人們常在取樣，樣本如何取得？該不該出來參選學生會會長？難以決定。問問幾個死黨的意見，被問的人，就是你取的樣。在眾人一致鼓勵下，充滿信心地報名參選，不料以慘敗收場。這才知眾好友都太友善了。想去某家餐廳，評價如何？上網看看。好評不少，去後卻大失所望。後來想通了，這種平價餐廳，顧客及會上網寫評論者，皆以年輕人居多。年輕人的愛好，自然跟你這位老先生不太一

樣。有些事無關緊要，可就取樣，以為參考。但對某新產品的市場接受度，可不能隨意找些人來問意見。

底下來看一道統計試題。104 學年的學測，數學科有如下一多選題：

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。

答案是(2)、(4)、(5)。“大考中心”很客氣，說所提供的僅是“參考答案”，而非“標準答案”。

首先，來檢視一下 5 個選項的敘述。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此次路跑的推論。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，這是疏失。另外，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，會讓人以為是針對一般情況提問，而不僅是此次路跑。所以，若依現有題目之敘述，有學生遵循邏輯，謹慎地未選(1)、(3)、(4)及(5)，應該算是對的。命題者對題目之敘述，似不夠謹慎。

再給一文字方面的問題。兩個隨機變數，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)，並不太通。宜問相關係數是否為正(或負)。

有人可能好奇，選項(4)(或(5))的敘述，可否改為“每公里步數和每公里平均心率為正相關(或負相關)”？即刪除“的相關係數”5 字。若這樣改，題目敘述便無瑕疵。但這只是一次路跑的數據，若小明再跑一次，或繼續跑 10 公里，可能得到完全迥異的數據，因而計算出之相關係數，連正負說不定都會反過來。這有如題目若先說“投擲一銅板 10 次，得到 5 個正面”，則“銅板出現正面的機率為 0.5”之選項，便不該選。因此實際上“每公里步數”，與“每公里平均心率”，此二變數是否為正相關，並無法由小明跑 10 公里後，產生的數據得知。所以，一旦敘述如上修改，則選項(4)便不正確。但既然參考答案中有(4)，表示命題者認為選項(4)是可判定為正確的，因此敘述便不可如此修改。至於選項(3)，由於未列進大考中心提供的參考答案中，因此若要刪除“的相關係數”那 5 字，自然是可以的。

在學測如此大型的考試，命題者對文字的陳述隨意，尚非本題最關鍵的缺失。要知就算題目寫得不清不楚，也大致能猜出命題者的意思，這是我們的中學生，早就被訓練出來的本領。假設某人想觀察自己體重的變化，每天記錄。能否想到該注意些什麼？須在相同的情況下記錄。例如，每天皆在剛起床時量，這樣較能相比。即使如此，每天起床時間可能有差，或前晚因應酬吃喝較多，就算採取剛起床時量測，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，連續跑 10 公里，並不易維持每 1 公里的狀

況都相同。因此，少有以這種方式，收集個人的數據。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。無論如何，用一個人自身幾個變數的數據，來考慮相關係數，意義並不太大。如應力與應變有線性關係，此為力學裡的虎克定律(Hooke's law)。小明個人的那些變數，甚至可能並不隨機，彼此之間，也可能僅有函數關係。

資料的收集，是從事統計工作，一很重要的步驟。惟有秉持很嚴謹的態度，取得的數據，方能準確客觀，因而得到的推論，才較具參考價值。就如醫學上，一種新藥，或新技術，其效果如何？需找人做實驗，也非徵求自願，來者不拒。不但要謹慎挑選受測樣本，且過程有一定規範，如須雙盲實驗。對某政治議題，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。取樣須很謹慎。

最後，對於上述那道考題，是否可假設小明每一公里，都維持相同的狀態？如果做這樣的假設，則那便是數學而非統計題目了。附帶一提，106 學年學測數學科，有一道用某城市 12 個月裡，各月的最低溫與最高溫，來考慮正負相關的多選題，也犯了與本題類似的缺失，讀者不妨找來看看。可以這麼說，如果連學測數學的命題者，對統計的理解都普遍不夠正確，則一葉知秋，高中數學的統計教學，是很令人擔憂的。

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。

5 估計

2017 年 1 月 12 日，聯合報有則標題是“到底幾歲？ 月球比原先以為的更高齡”之報導：

> 根據 1971 年阿波羅 14 號太空船採集到的岩石和土壤，最新估算出來的月亮年紀，比先前許多科學家設想的更大：已高齡 45.1 億歲。
>
> 洛杉磯加州大學(UCLA)的研究團隊 11 日表示，月球在太陽系生成後的 6 千萬年間形成。先前的估算落在太陽系出現後 1 億年間，最晚的是 2 億年。…。月球由地球被碰撞後的殘留碎片所形成，科學家估算地球約 45.4 億歲。巴邦尼說，她正在研究更多來自阿波羅 14 號樣本的鋯石，但不認為會改變她的估算—月球約 45.1 億歲，最多 45.2 億歲。他們的報告 11 日刊登在“科學進展”期刊上。

由一些石頭及土，便能對距地球約 38 萬公里外的月球，估計其年齡，相當厲害。將之前的估計，月球形成於約 45 億年前，修訂成 45.1 億年，也就是肯定早先估計的 45 億年是可靠的，只少了 0.1 億年。此報導除顯示，估計難免有誤差外，亦可得知，什麼都可拿來估計，包括太空裡星球的年齡。

棒球比賽，緊要關頭，站上打擊位置的球員，會不會成為致勝功臣？他打出安打的機率約為 0.22，機會看來不是很大。0.22 是如何得到的？由該球員之前的打擊率來估計。看起來合理，機率裡稱此為頻率的觀點。若有人覺得該球員近況很好，過去 1 個月，平均打擊率有 0.43，因此估計他這次擊出安打的機率為 0.43，亦無不可。有人出門前，估計今天下雨的機率為 0.8。如何得到？看看天色，腦海中就出現 0.8 這個數字。喔！這是主觀的觀點。只是同樣的天色，不同的人，估計出的下雨機率，可差異很大。前往高鐵站的路上，遇到塞車，趕得上預訂的班次嗎？這支股票會漲嗎？這西瓜甜嗎？生活上人們經常在估計。估計的確該是國民基本的統計素養。無關緊要的事，憑一些現有資料，用些簡單的統計，便能做出估計。有時就要妥善規畫一套估計的程序，以減小誤差。

不但可以有不同的估計法，面對一樣的資料，也可有不同的推論。在黃文璋(2011a)一文裡，我們提到“事實至於遺存，推論敬俟卓識”，這是考古學家郭寶鈞(1893−1971)所說的，他曾參與多次殷墟挖掘。統計與考古，本質上很類似。考古有物而無言，挖掘固然困難，但目的是給出推論，為古物說故事。統計裡數據取得不易，而一旦蒐集到，數據並未說話，整理分析後，由人負責說話，給出推論。至於那一推論較佳？這便難說了。自一有紅球及白球的袋子中，依序取 20 球，每次取出後皆放回，結果 19 次得白球 1 次得紅球。雖取出之白球比率遠高於紅球，有人仍認為下次取，將會得紅球。頑固嗎？說不定還真的取出紅球。因此對隨機現象的

估計或推論，不能以一次的準或不準，來分高下。得給出評比的準則，才能定下各估計法或推論法之優劣。

前面提到的皆是所謂點估計，即以一值，來估計一未知的量。某日小明考完數學，媽媽問考得如何？大約 8、90 吧！有幾成把握？9 成。這是典型母子間的對話。由於不太確定閱卷標準，對成績的估計，常不會只給個值，而給個範圍，並附上信心大小。對一隨機現象，估計值並不易精準命中真實值。於是除點估計外，發展出以一區間來估計一未知的量，並附上該量會落在此區間之機率。其中的區間，稱為信賴區間，伴隨的機率，則稱為信心水準。在前述小明的成績估計裡，信賴區間即[80,90]，信心水準則為 90%。

有幾個問題立即會浮現。首先，信賴區間唯一嗎？在估計一未知量時，若採區間估計，常是先給定信心水準，然後求出信賴區間，使該區間涵蓋未知量之機率，如事先所設定的信心水準。信心水準一般以百分比表示，如 95%及 90%等，皆是常取的信心水準。對一固定的信心水準，信賴區間往往不唯一，一段採取區間長度較短的。區間的長度愈短，表估計愈精準。其次，信心水準的涵義為何？95%是代表機率嗎？取樣前信賴區間是一隨機區間，說有 95%的機率，包含待估計的參數，這點沒有疑義。取樣後則得一常數區間，此區間要嘛包含參數，要嘛不包含，如何說包含的機率是 95%？之前在高中數學裡，碰到的機率問題，大抵與排列組合有關，以相同的可能性來解釋機率，那時師生較少去想機率究竟是什麼。但自引進信賴區間後，不少人連機率的意義，也開始

感到疑惑了。事實上，只要是未知便能談機率。正如考前教師早已出好題了，學生仍可猜題，還說這題考的機率有 8 成，那題不可能考等。既然取樣後所得的固定區間，可以對它談機率，則利用主觀機率的想法，將信心水準，視為該區間包含欲估計參數的機率，乃再自然不過。

我們說過大學生對信賴區間的涵義，大多僅一知半解。那何以少聞在大學統計學的課程裡，對信賴區間之學習有何困難？考試引導教學，我們不妨來看一下高中數學裡，信賴區間都考些什麼？

底下為 98 學年，學測數學的一道多選題：

> 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]。試問下列哪些選項是正確的？
>
> (1) 甲地本次的參訪者中，54%的人聽過該產品。
> (2) 此次民調在乙地的參訪人數少於在甲地的參訪人數。
> (3) 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%。
> (4) 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]。

(5) 經密集廣告宣傳後，在乙地再次進行民調，並增加參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”公佈的答案為選項(1)及(2)。

選項(3)不在答案裡，此很值得斟酌。我們已說了“推論敬俟卓識”，但誰又是卓識？依其理念之不同，統計裡可有各種推論，及各種解讀。被視為一意孤行者，還可能自以為擇善固執呢！但在數學裡就不行，有一套邏輯，不可胡亂推論，隨意解讀。選項(3)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？

2017 年 1 月 11 日，大陸遼寧艦(即遼寧號航空母艦)，從海南島返回青島駐地的途中，穿越台灣海峽，引起多方的重視。有人認為是繼 1996 年大陸試射飛彈以來，對台灣最具挑釁意味的軍事行動；也有人指出，遼寧艦事實上還不具備完整戰力，但穿越台海在台灣內部引發的討論，已經達到中共宣傳戰設定的目標；有人則相信這不過是大陸一個例行性

的軍事演習，不必太在意。對此事件，學者專家各有其解讀，，惟真相並不易大白。要知解讀與推論，可以相當主觀，很難判定何者正確。因此考題若以“可解讀為…”，或“可得推論…”，的形式提問，往往缺乏統計的味道。

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 $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	0.02	0.04

請問從此次抽樣結果可以得到下列哪些推論？

(1) 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。
(2) 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)。
(3) 此次抽樣的女性公民數少於男性公民數。
(4) 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。
(5) 如果不區分性別，此次抽樣 $\hat{p}$ 的標準差

$\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ 介於 0.02 與 0.04 之間。

“大考中心”公佈的答案為選項(2) ，及(4)。

選項(2)、(3)、(4)，及(5)，乃考是否熟悉課本上所給信賴區間的公式。(3)、(4)，及(5)，甚至根本是在考數學。至於選項(1)，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(1)不被命題者視為正確選項的原因。但比較二未知比例的大小，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(1)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。

在黃文璋(2011b)一文，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。想以區間來估計，便該多求出信賴區間。大學的統計課裡，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，有相當多的參數可估計。於不同的情況下，探討如何求得信賴區間。由於包含的內容不少，因此進度慢不下來，學生匆匆忙忙學習，囫圇吞棗，根本無暇去思考信賴區間的內涵。這是何以多年來，大學信賴區間的教學，少見困擾的主因。並非學生皆能融會貫通，而是問題沒有浮現。但在高中數學裡，就只針對二項分佈 $B(n,p)$，要求參數 p 的信賴區間。以常態分佈來近似，且信心水準設定為 95%。內容少，因而求信賴區間的題目，便缺乏變化，加上計算又容易，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，有些教授硬著頭皮去做答，卻不乏 5 個選項中，錯了好幾個的。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”要正確才易答對，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 105 學年，4 年間的學測及指考，及 106 學年的學測，共 13 份數學試題中，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。

有關估計及信賴區間相關的文章，可參考黃文璋(2006b)、(2007a)，及(2014c)三文。

6 極限

如同故事結局引人關注，對一進行中的事，人們屢會好奇，一直下去最後會怎樣？這其實就是想知道極限為何，極限是數學中一重要的概念。大學裡很多學系的必修課中，均列有微積分。因這是進入不少領域的敲門磚，而微積分的基礎便是極限。由於高中的選修數學裡，已有微積分的題材，所以極限對一般高中畢業生，並不陌生。

極限是什麼？簡單講，有一數列 a_n，$n\geq 1$，考慮 n 趨近至無限大(以 $n\to\infty$ 表之)時，a_n 會趨近至何值？或有一函數 $f(t)$，$t\geq 0$，考慮 $t\to\infty$ 時，$f(t)$會趨近至何值？這是微積分裡，兩個求極限之基本形式。至於怎樣是趨近？那可非三言兩語能解釋得清楚。有些人還誤以為趨近乃接近之意。像曾在一則新聞裡出現"至於籃球，相較起來，兩隊總得分出現單或雙的機率，就比較趨近於相等"一句。其中的"趨近於相等"便非正確敘述，宜以"接近"取代。又極限當然也可能不存在，極限不存在的討論，常會讓學生搞得糊裡糊塗。換句話說，要真正弄懂極限，並不十分容易。即使在大學裡，修過微積分的人，對極限多半也只是懵懵懂懂，不敢多談。雖然如此，機率與統計裡，兩個基於極限的隨機法則，即中央極限定理與大數法則，不少人卻能朗朗上口，隨時引用，一無所懼。

2014 年 11 月底九合一選舉投票前，有位臺北市長候選

人，拋出以遴選委員會，來挑選首長的主張。他解釋：

> 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”。

N 是什麼沒說，且 $N>25$ 時，何者會接近大數法則也沒說。只是大數法則，顧名思義是個法則(law)，引用無妨，至於說“接近它”，就不知意義為何？另外，大數法則與中央極限定理(通常稱“定理”而不稱“理論”)內涵完全不同，所以沒有“大數法則也就是中央極限定理”這種講法。而大數法則居然能用來遴選市政府首長，還認為蠻準確，更是聞所未聞。

這位候選人究竟在說些什麼，一般人可能並不太清楚。但看他一付言之鑿鑿，即使對其講法心存懷疑者，恐怕大多數寧可不開口了，以免自曝其短。畢竟知之為知，不知為不知，是知也。

先來看大數法則，這可能是較容易理解的。氣象局每日公佈對各地降雨機率之預測。只是就算預告的降雨機率高達90%，一天結束，也不見得果真下雨。氣象局的預測準確嗎？令人好奇。你知道不能僅由幾天的準不準，便驟下結論。就像對一銅板，若只投擲幾次，並無法判定銅板是否公正。我國中央氣象局，在其網站上，對“降雨機率 60%”的說明為，每預測 100 次，實際降雨有 58.6 次。每 100 次，怎會有不是

整數的 58.6 次？這應是指“平均”而言，也就是若過去預測“降雨機率 60%”的有 n 次，其中 k 次真的下雨，則 k/n 約為 0.586。由於氣象局已長期進行預測，這裡的 n 當然不會小。氣象局所提供對機率的解釋，乃符合常見頻率的觀點。

一隨機試驗若有兩個結果，即稱為伯努力試驗(Bernoulli trial)。這種試驗，雖簡單卻處處出現。如投擲一銅板，會出現正面或反面；投擲一骰子，會出現偶數或奇數等。一般而言，若對一隨機試驗的某事件 A 有興趣，則觀測後，可能事件 A 發生，也可能事件 A 沒發生，如此便是兩個結果了。對於二結果，人們習於其一以成功稱之，其二以失敗稱之。成功常只是表有興趣的那一結果，並不見得真的做什麼事成功。

假設有一可重覆觀測的伯努力試驗，每次成功的機率以 p 表之。則每次試驗後，便產生一有伯努力分佈之隨機變數，即取值 0 或 1，其中 1 表成功，0 表失敗，且得到 1 的機率為 p，得到 0 的機率為 $1\text{-}p$。觀測 n 次後，以 S_n 表總共成功的次數，則 S_n 有二項分佈 $B(n,p)$。而 S_n/n 表觀測 n 次後，平均每次觀測成功的次數。S_n/n 即為 n 次試驗後，所得 n 個伯努力隨機變數之樣本平均。不論專業或非專業人士，皆知可以 S_n/n 來估計 p。統計學裡，則告訴我們，此一 p 之估計量，有很多好的性質。

當觀測次數 n 不太大時，S_n/n 取值還可能上下有不小的波動，但當 n 很大時，S_n/n 便“大致”會很接近 p。即若在座標平面上，繪 $y=S_n/n$，$n\geq 1$，之圖形，則隨著 n 之愈來愈大

後，其圖形將“大致”貼近水平線 $y=p$。這便是一般人以為的大數法則，即使沒正式學過，直觀上也覺得會成立。之前所提以頻率的觀點來解釋機率，其背後的理論基礎，便是大數法則。對大數法則，有時會將其詮釋成，一件事最終會呈現該有的風貌。所指的就是上述觀測次數夠多後，樣本平均將“差不多”等於事件發生機率的意義。所以，想估計銅板出現正面的機率 p，人們向來知道，就是持續投擲，只要投擲數夠多，則以總共出現的正面數除以投擲數，便是 p 之一不錯的估計值。

要注意的是，前述討論其實隱含做了一假設。即各次投擲相互獨立，且每次出現正面的機率相同。也就是假設各樣本為相互獨立且有相同分佈。因此不會有這次出現正面，下次便較易出現正面；或投擲若干次後，出現正面的機率改變等情況。實務上，如果以隨機取樣，來估計池中某種魚之比例，或市民對某議題之支持率，由於魚與人皆有“個性”等因素，此假設並非毫無疑義地成立。但對於投擲銅板，此假設通常可接受。只是不論 n 再大，S_n/n 不但不見得會等於 p，甚至也不保證就很接近 p。以投擲一公正銅板 100 次為例，是有可能全出現正面，雖發生的機率$(1/2)^{100}$ 相當小。此時，$n=100$，$S_{100}=100$，因而 $S_{100}/100=1$，可一點都不接近 $p=1/2$。

我們已指出，對 $n=1$，2，…，隨機數列 S_n/n，隨著 n 之增大，並不必然很接近 p。但這豈非違反大數法則？倒也沒有。那大數法則是在講些什麼？以白話來說，乃表只要 n 夠大，S_n/n 任意接近 p 的機率，便可任意接近 1。而怎樣算是

任意接近？即差距可以小於任給的一個上限，所給的上限當然須是正的。以數學來表示，對任給的二正數 a，b，其中考慮 $b<1$ 即可，則只要 n 夠大，便能使$|P(|S_n/n-p|\leq a)-1|<b$。又因機率不會大於 1，前述不等式等價於 $P(|S_n/n-p|\leq a)>1-b$。所以，只要 n 夠大，S_n/n 雖不一定很接近 p，但會很接近 p 的機率，便將很大，要多大皆可。我們之前提到的“大致”，便是這個意思。而由於是在談機率，當 n 很大後，S_n/n 不接近 p 的機率既然很小，要多小皆可，那就夠了，不必在意 S_n/n 不接近 p 的情況。

大數法則告訴我們，對任一給的正數 a，只要 n 夠大，則 S_n/n 與 p 之差距不超過 a 的機率，將差不多是 1。討論隨機現象，機率是 1 或 0 的事件，表必然發生或必然不發生，這樣的事件，通常不會讓人太感興趣。由於機率差不多是 1，顯示若想知道 S_n/n 究竟如何散佈在 p 附近，不論 a 如何小，當 n 持續增大後，S_n/n 與 p 之差距不超過 a，便遲早是一“太大”的範圍(表 S_n/n 落在此範圍內之機率差不多是 1)。因此欲較“有意義的”描述 S_n/n 散佈在 p 附近的情況，所給的 S_n/n 落在 p 附近之範圍，須更小才行。固定的範圍，已知不適用了。最後求出，適當的範圍，其尺度單位為 $1/n^{1/2}$，乃一隨著 n 之無止盡地增大，而漸減至 0 之尺度。如此一來，便有辦法來近似 S_n/n 落在 p 附近一範圍內之機率。更明確地說，對 $\alpha<\beta$，n 很大時，

$$P(S_n/n\in[p+\alpha p(1-p)/n^{1/2}，p+\beta p(1-p)/n^{1/2}])，$$

即$(S_n/n - p)/(p(1-p)/n^{1/2})$落在區間$[\alpha,\beta]$的機率，即

(1) $$P(\alpha\leq(S_n/n-p)/(p(1-p)/n^{1/2})\leq\beta)$$

$$=P(\alpha\leq(S_n-np))/(np(1-p))^{1/2}\leq\beta),$$

上式左側是關於樣本平均 S_n/n，右側則是關於樣本和 S_n，可以函數

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

由 α 至 β 的積分來近似。此結果，便是著名的二項分佈之常態近似，也是很原始版本的中央極限定理。

(2)式為標準常態分佈(以 $N(0,1)$表之)之機率密度函數，其中包含數學中幾個重要的常數、函數及運算。以常數來說，有最小的正整數 1、最小的偶數 2、圓周率 π、自然對數函數的底 e。由於 S_n 的期望值為 np，標準差為$(np(1-p))^{1/2}$，減去期望值後，再除以標準差，便是將 S_n 標準化。中央極限定理即指出，樣本和 S_n 經標準化後，當 n 很大時，其分佈可以標準常態分佈來近似。函數 $y=\phi(x)$，$x\in R$，其圖形有如鐘形，對稱於 y 軸，最高點發生在 $x=0$，自 y 軸向兩側漸減至 0。

機率裡分佈的產生，有一些不同的途徑，經由極限為主要的一種。常態分佈便是極限下，所產生的一重要無比之分佈。

大數法則說，n 很大時，S_n/n“大致”在 p 附近。但“大致”一詞乃口語，若想知道如何大致法，則大數法則就束手無策了，得換由中央極限定理來發揮功能。中央極限定理能給出，S_n/n 落在 p 附近一範圍內之機率的近似值。而若不欲

機率值“太大”，範圍便得隨著 n 之增大，而愈來愈小。中央極限定理，相當於用倍數以 $n^{1/2}$ 的速率成長之顯微鏡，來觀看 S_n/n 散佈在 p 附近之情況。若要讓散佈情況看起來更優美些，便將倍數調整為 $n^{1/2}/(p(1-p))$。

投擲一公正的銅板 100 次，以 X 表正面出現的次數。大數法則指出，$X/100$“大致”是 0.5。即有很大的機率，$X/100$ 會很接近 0.5。那是否表 X“大致”是 50？如果答案為肯定，由於 X 取整數值，因此 X 便有很大的機率為 50 了。利用排列組合，會出現 50 個正面的機率為

$$P(X=50)=\mathrm{C}(100,50)(1/2)^{100}=100!/(50!)^2\times(1/2)^{100}。$$

只是 100!很大，上述機率值要正確算出不易。藉由近似公式，可得

$$P(X=50)\approx 0.0796，$$

為一不算很大的機率。所以大數法則，並未使投擲一公正的銅板 100 次後，出現的正面數“大致”為 50。$X/100$“大致”是 0.5 沒錯，但小差距放大 100 倍，可不見得仍是小差距了。X 的散佈範圍變大了。事實上，投擲一公正的銅板 $2n$ 次，當 n 愈大，將愈不容易出現恰好 n 個正面。

X 之期望值為 50，且標準差為 5。底下藉求 X 落在期望值 50 附近一範圍內的機率，來看中央極限定理之一應用。先求 $P(45\le X\le 55)$，即 X 與期望值 50 相距不超過 1 個標準差之機率。利用(1)及(2)式，及 $N(0,1)$介於正負 1 間之機率，約

為 0.6826，即得

$$P(45 \le X \le 55) \approx 0.6826。$$

或者就簡單的說機率約為 0.68。至於 X 與期望值 50 相距不超過 2 個標準差之機率，仍由中央極限定理，得

$$P(40 \le X \le 60) \approx 0.9544。$$

或者就簡單的說機率約為 0.95。

有幾點要說明。首先，當 n 很大時，分佈可以 $N(0,1)$來近似的，既不是 S_n，也不是 S_n/n，而是標準化後的 S_n。包括若干高中數學課本的作者，很多人誤解這點。還宣稱隨著 n 之增大，S_n 及 S_n/n 的機率密度函數之圖形，都會愈來愈像 $N(0,1)$那一鐘形的機率密度函數圖形。其實很容易可看出這並不正確。S_n 不會取負值，且隨著 n 之持續漸增，S_n 取值很大的機率，將愈來愈接近 1；至於 S_n/n 取值乃介於 0 與 1 之間。因此就算 n 再怎麼大，不論 S_n 或 S_n/n，其機率密度函數的圖形，一點都不會近似定義在整個 x 軸，函數 $y=\phi(x)$，$x\in R$，那一對稱於 y 軸的圖形。這方面的詳細討論，可參考黃文璋(2011c)一文。

自問世以來，歷經不同學者的推廣，大數法則及中央極限定理，條件均早已放寬不少，因此適用性更廣。也讓諸如人的身高、體重、智商，及量測的誤差等，常可以常態分佈來當模式。常態分佈遂處處可見，於眾多分佈裡，享有獨尊的地位。在常見的版本裡，不限伯努力分佈，對一數列相互

獨立且有共同分佈的隨機變數，當共同分佈的期望值存在，大數法則便適用；當期望值及變異數皆存在，則中央極限定理便適用。雖相互獨立與分佈相同，皆可以較鬆的條件取代，但就是須有些條件。並非只要有很多隨機變數，大數法則或中央極限定理，便能派上用場，這點須特別留意。

回到本節一開始所引那位臺北市長候選人的談話。大數法則或中央極限定理，雖為極限下的結果，但實務上，遇到的樣本數自然都是有限。有些教科書指出，“通常”樣本數 n 並不必太大，如 n 至少達到 30，或者更小些，n 至少有 25，則關於樣本平均，及樣本和之機率，就都可利用中央極限定理來求近似值。這可能是那位候選人說“在統計學上，$N>25$ 就會接近大數法則，也就是中央極限定理”，其中“$N>25$”的由來。但此不求甚解的候選人，不但將大數法則與中央極限定理混為一談，甚至並不了解大數法則的內涵。只要樣本數夠大，就不太會出現偏頗嗎？不妨以電影獎項之評審為例。

世界各地年度的大小電影獎項很多，在台灣便設有金馬獎，及台北電影獎等兩個，因此全世界超過 100 個應毫無疑義。其中由美國影藝學院(Academy of Motion Picture Arts and Sciences) 所主辦的奧斯卡金像獎(The Academy Awards，或簡單稱之為 Oscars)，不僅是美國，甚至可說是全世界最受矚目的電影獎。此學院的會員總數，至 2016 年 12 月底，共有 6,687 人，是影藝界一很大的組織。每年頒發包含最佳影片、最佳導演、最佳男主角、最佳女主角、最佳男配角，及最佳女配

角等，共二十餘個獎項。各獎如何產生？先由相關分會的成員，投票決定入圍名單。如演員會員(有 1 千 3 百多位)，可對最佳男、女主角，及最佳男、女配角獎，有關演技的 4 個獎項投票。待全部獎項的入圍名單都確定後，每一會員，皆可對任一獎項投票。投票人這麼多，又都是專業人士，應不會有爭議吧!

聽過“奧斯卡好白”(Oscars So White)嗎？2016 年 1 月，奧斯卡金像獎入圍名單公佈後，掀起的“種族歧視”質疑聲浪，幾乎淹沒了提名的喜悅。因連續兩年，有關演員的 4 個獎項，20 位入圍者(4 獎每年均各有 5 人入圍)，清一色都是白人。此外，幾個重要的獎項，如最佳影片，及最佳導演等，也幾乎是白人的天下。每一獎項，參與入圍名單投票的人數，均遠遠超過 25，甚至上千，但因此就不會出現偏頗的情況嗎？由影藝界的反應來看，顯然他們極不同意。隔年，2017 年，情況大改變，有 7 名非裔演員，入圍奧斯卡 4 個演技獎。20 分之 7，所佔比例相當高，且是奧斯卡史上，首次所有演員類別獎項，都有非裔入圍。其中最佳女配角獎，5 位入圍者裡，便有 3 位非裔。而結果揭曉後，獲獎的最佳男、女配角，皆為非裔。是抗議有效，還是入圍與否，僅依演員表現，而沒有其他考量?讓人有不少的遐想。只是雖黑白協調了，仍有人覺得美中不足，因亞裔等少數族群，露臉的程度還是少。話說回來，若真的做到對各族群面面俱到，恐怕獲獎名單的產生，是經特別安排，而非全憑表現了。

一般在做民調時，會在意誤差的大小。直觀上，欲誤差

小，樣本數便不能少。為使誤差不超過 3%的機率，大於 0.95，藉助中央極限定理，同時也採用一些其他的近似(見黃文璋(2011d))，一開始設定的成功樣本數為 1,068。1 千個左右，宣稱是隨機所產生的樣本，由於有拒訪等人為因素，都還常被批評取樣的代表性不足，因此做出的調查偏差太大，未能準確反映真相等。為特定目的而成立的遴選委員會，了不起數十個委員，怎敢說一定客觀？其委員幾乎不可能隨機產生，大抵是被指派。因此從委員會成立之始，便難以避免主觀了。而委員不乏具同質性或相異立場，既難以獨立，也不會有“共同分佈”。因此不論大數法則或中央極限定理，在此皆不適用。怎會異想天開，拿中央極限定理，來為遴選的公正性背書？

2014 年 11 月落幕的第 51 屆金馬獎，得獎名單由總共 17 位委員共同討論決定。但未獲最佳女主角獎的著名演員鞏俐，事後透過經紀人表示，金馬獎不專業、不公正，今後不會再參加了。由此可看出，像遴選這類評比，有相當程度的主觀成分，爭議難免，並無所謂“只要 15 個人就會蠻準確的”這種推論。

大數法則及中央極限定理，乃機率中二極重要的隨機法則，用途廣泛，但卻也沒那麼無所不包，連選才都用的上。正確運用為上，不必過度推崇，也不可意圖用來唬人。

數學裡常在證明，在給定的前提下，推導出某些結果成立。如何推導？如同下棋、打牌，各有規則，數學中亦自有一套演繹的邏輯。由給定的前提，一步一步完成證明。不擅長證明的人，數學通常也就學不好。棋下不好，或牌打不好，對人生可能沒什麼影響。但對中學生，若數學不佳，可能會影響到能進的大學學系。幸好就只是這樣而已，不擅長數學，對人生倒不見得有什麼大不好。

以命題來描述，給定的前提稱為 A，欲推導出來的結果稱為 B。若 A 成立，則 B 必成立，便稱此命題為真(或稱成立)，否則稱此命題為偽(或稱不真、不成立)。數學裡，便經常在證明一個又一個的命題為真或為偽。某命題一旦被證明為真，除非能找出漏洞，推翻證明，否則便不必試圖去找反例。只是世事以真偽難辨的居多。由於無法證明，科學家遂習於以“證實”一詞，宣佈其發現。以咖啡為例。“研究證實：愛喝咖啡助長壽”，這是一則新聞的標題。事實上，不時有研究報告指出，喝咖啡的好處多多，除有助長壽外，亦有能降低罹患阿茲海默症的風險之說法。當然證實喝咖啡有害人體的報告也不少。但畢竟僅證實並非證明，因此常是信者恆信，不信者恆不信。科學上究竟是如何證實呢？

“設有一公正銅板，投擲 20 次，試求出現 19 個正面的機率。”這是高中數學裡，一道典型的題目。公正銅板，說起來輕鬆無比。因數學中的假設是不須證明的，所以不必質疑銅板是否真的公正；或去想世界上是否存在公正的銅板？而且眾所皆知，做這種題目，並不需去投擲。甚至，根本不

必有銅板，就能求出答案了。不過，若真拿到一銅板，且投擲 20 次，出現 10 個正面，自然不會懷疑銅板的公正性；但如果出現 19 個正面呢？恐怕將難相信銅板為公正。只是不論怎麼投擲，都無法證明銅板公正或不公正。銅板是否公正，永遠只有天曉得。實務上，銅板公正，僅能是個假設，就看接受或不接受。如何判定？也就是怎樣才會接受銅板為公正？投擲 20 次，若只有正面恰好出現 10 個，才相信銅板公正，便未免太嚴格了。因即使銅板真的公正，此機率才約 0.1762。如此一來，將有約 82.38%的公正銅板，被誤判為不公正。統計裡發展出檢定一假設，是否可接受的程序，即假設檢定(testing hypothesis)。此亦顯示統計與數學的不同。之前說，數學是從給定的前提下，推導出結果。統計則是由給定的結果，來驗證前提為何。

很多人都聽過檢定，有各種檢定。如象棋與圍棋的棋力檢定等。參加檢定者，依據其測驗或比賽的成績，是否達到某一事先設定的標準，以判定能否通過某級或某段之檢定。想知道自己棋力者，會去參加檢定？而原本已擁有某級(段)證書者，近來有時打敗比他高一級(段)者，自覺棋力進步了，想要晉級(段)，也會去參加檢定。但由於臨場反應等因素，其中無法避免會有運氣成分。

如前所述，數學裡的命題，一旦被證出，就毫無例外地成立，由不得你接不接受，因那已是一事實。統計裡針對一有興趣的假設，若通過檢定，便接受此假設。而就像棋力檢定，檢定的通過與否，機運免不了。因此一通過檢定，被接

受的假設，並不表就是事實。這說明了科學上所宣佈的證實，何以總會有人不相信。因那些所謂證實，往往是基於通過了統計假設檢定。故若有人以為通過檢定，只是運氣好，因而不肯相信，乃可理解。科學上的研究，涵蓋面很廣，各式各樣的證實很多。對諸如“科學證實快樂的人有這 10 個共同點”之類的報導，目不暇給之餘，或許會讓人產生疑惑：就算號稱通過統計檢定，是否假統計之名？

在假設檢定的架構裡，有虛無假設(null hypothesis)，與對立假設(alternative hypothesis)二假設。虛無假設通常表現況，或傾向想推翻的；而對立假設則表傾向接受的。與刑法上的無罪推定原則一樣，雖目的是想推翻虛無假設，但作法卻是儘量保護它，除非佐證夠強，否則不會推翻虛無假設。如此一旦推翻，接受想要的對立假設，才較具說服力。若與棋力檢定的例子對照，虛無假設即目前擁有證書上的級(段)數，對立假設則相當於擬晉的級(段)數。檢定的結論將是接受虛無假設未晉級(段)，或接受對立假設晉級(段)。

之所以執行一假設檢定，乃是因懷疑虛無假設不真。由此應能明白虛無假設何以名之為虛無了。虛無(null)就是空，若經過檢定後，接受虛無假設，表整個過程白費，退回到現況。想想消費者保護會懷疑某食品有毒，大費周章地進行檢定後，宣佈食品合格；被檢察官起訴的嫌犯，最後被法官判定無罪，這些都將落入天下本無事，庸人自擾之的窘境。

在假設檢定裡，通常以 H_0 及 H_a，分別表虛無假設及對立假設。舉例來看。令 p 表某銅板出現正面的機率。原本銅

板被視為公正，若懷疑銅板較易出現正面，則取 H_0：p=1/2，H_a：p>1/2；若明確地懷疑 p 應是 0.55，則取 H_0：p=1/2，H_a：p=0.55；若只是覺得銅板不公正，則取 H_0：p=1/2，H_a：$p \neq 1/2$。對隨機現象做決策，要從不誤判幾乎是不可能。會有兩種可能的誤判。其一是 H_0 為真，卻拒絕 H_0，接受 H_a，稱此為第一型錯誤；其二是 H_a 為真卻拒絕 H_a，接受 H_0，稱此為第二型錯誤。在無罪推定之原則下，通常犯第一型錯誤比較嚴重。因食品明明合格卻被判定有毒，則商譽及銷售都將受損；根本無罪卻被判有罪，可能受罰或坐牢，甚至毀了人生，這種錯誤較難彌補。

先設定一個第一型錯誤機率值之上限 α，稱為顯著水準(significance level)。通常 α 為一較小的值，0.05，0.01，及 0.001 等，都是常取的值。實務上，究竟怎樣的機率才算小，當然須視情況而定，不能一概而論。如果嫌犯被以死刑罪起訴，則 α 取成百萬分之 1 都不算小。在 α 給定後，便決定何時拒絕 H_0，即決定拒絕域。在 H_0 為真的假設下，出現的結果，若落在拒絕域(機率不能超過 α)，便稱得到顯著的(significant)結果，或者說檢定結果為顯著。怎樣的事件才顯著？人們常說“狗咬人不稀奇，人咬狗才稀奇”。小機率事件才屬顯著，因這種事件若發生，將讓人不可等閒視之。結果顯著，動搖了虛無假設的基礎，便該拒絕 H_0。出現的結果，若沒有落在拒絕域，表所發生乃機率大的事件，也就是觀測到稀鬆平常的結果，不足為奇。既然 H_0 的假設，沒有使不合理的現象發生，那就宜照舊，即仍接受 H_0。執行一檢定，暗中希望的，是得到顯著的結果。結果顯著，拒絕 H_0 接受 H_a，才表

檢定沒白做，是一成功的實驗。在同一 α 下，拒絕域的選取，若能使第二型錯誤的機率值 β 最小，當然最好。此時的拒絕域，稱為顯著水準不超過 α 下之最佳拒絕域。在某些條件下，統計學裡有一套找到最佳拒絕域的方法。

既然犯第一型錯誤較嚴重，那 α 是否取得愈小愈好？一般而言，α 愈小 β 將愈大，無法兩全。以法庭審判為例，若法官悲天憫人，寧可錯放 1 千，而不願錯罰 1 人，對證據的審核高度嚴格，則很多實際有罪者，將被宣判無罪，興高采烈地回家了；作假的不肖商家，可繼續矇騙顧客了。所以 α 取得過小，便太保護虛無假設，明明不真卻拒絕不了，並非好事。但也不宜為了使檢定易成功，而取太大的 α，這樣現況將常被推翻。須視不同狀況，斟酌取適當大小的 α。

某公司欲檢定某種研發出來的飲料，喝了是否有助減肥。取 H_0：減肥無效，H_a：減肥有效，且取 α=0.05。則即使該飲料完全無助於減肥，仍有 20 分之 1 的機率，會得到顯著的結果，因而接受設該飲料對減肥有助益。除了那 20 分之 1 的犯錯機率外，會不會有人持續做實驗，最後只公佈結果顯著的那件？也就是假統計之名。說不定就是因懷疑這種可能性，使有些人抗拒採信科學上的證實。不誠實的人自然存在，但夜路走多了，總會遇到鬼，作假遲早會被發現。執行假設檢定，萬不可選擇性的公佈結果。

曾修習統計學的課程者，多半以為假設檢定屬較難的題材，至少比信賴區間難。之所以覺得困難，大抵是找最佳拒絕域的部分。不過在很多情況下，憑直觀所得的拒絕域，常

就是最佳拒絕域。換句話說，實際應用時，即使沒去想拒絕域是否最佳，但往往得到的就是最佳拒絕域。例如，欲檢定 H_0：p=1/2，H_a：p>1/2，其中 p 表某銅板出現正面的機率。該如何進行檢定呢？要有數據，因此當然是拿起銅板投擲了。投擲 n 次，以 X 表正面出現的次數。則直觀上，會將拒絕域，取成$\{X\geq c\}$的形式。即若正面出現較多次，便接受 H_a：p>1/2。至於如何選取 c？乃由 n 及 α 決定。此一自動浮現出來的，就是最佳拒絕域。

我們再介紹 p-值(p-value)。進行一檢定，當得到一觀測值後，人們常會給出 p-值。所謂 p-值，即在 H_0 下，會得到比觀測值，至少同樣極端的事件之機率值。對前述檢定銅板是否公正之例，設 n=100，且觀測到 X=63。則 p-值便為 p=1/2 時，事件$\{X\geq 63\}$之機率。投擲一公正銅板 100 次，所得正面數比期望值 50 至少大 13，可求出機率約為 0.0062。得到此 p-值後便知，只要給定的 α 比 0.0062 小，就不能拒絕 H_0；而若給定的 α 大於或等於 0.0062，就得拒絕 H_0。給出 p-值，將較能看出，是很有把握地(p-值比 α 小很多)拒絕 H_0，或勉強地(p-值僅比 α 略小)拒絕 H_0。

空氣品質有沒有變好？石化工業對健康是否有影響？這類備受各界矚目的問題，其判定自然得依據嚴謹的假設檢定。至於 13 號星期五較不吉利嗎？此君值得信賴嗎？這些我們生活上經常會面臨的抉擇，與其靠占卜、星座等，其實亦該仰賴假設檢定的思維。假設檢定是為做較佳決策，該具備的統計概念。

黃文璋(2016) 為

8 檢定誤差

在隨機世界裡，常無法證明真偽，但為了做決策，不時得判定，或給出推論，這時假設檢定往往能派上用場。由於針對隨機現象，不論再好的方法，總免不了有誤差。假設檢定裡的兩型錯誤機率，便都是用來表示誤差大小。必須一提的是，藉助假設檢定做判斷時，得儘量保持客觀的態度，從多方檢視可能會有的誤差。統計畢竟只是用來協助做決策的工具，不必過度“推崇”。如果盲目的假統計之名，則犯下的錯誤，其嚴重性可能遠超過前述兩型。待犯下大錯後，倒過來抱怨統計不可信，便將多犯一個冤枉統計的錯。

宋朝的包拯(999–1062)，以清廉公正聞名於世，後世稱為他“包青天”，或“包公”。民間傳說中，有很多關於他破解各種奇案的故事。其中有一則常掛他名字的“包公審錢案”。賣油條的小孩，才離開一陣子，回來後發現錢不見了。原本錢放在盛油條的籃子裡，籃子則擺在一塊大石上。包公接獲報告後，叫人把石頭抬來審問。雖一再恫嚇，石頭不說就是不說，這麼頑劣？只好用刑了。只是打到棍子斷了，石頭仍不開口。一旁看熱鬧的人，都忍不住笑起來。包公生氣了，罰圍觀者每人拿兩個銅錢，扔進一個裝一半水的盆子裡。有個人才扔完錢，包公便指著他厲聲說“別走！就是你！”那人大呼冤枉，眾人也不解。包公說，“籃子裡的錢

都沾著油，而只有你的錢扔進水裡後，有油浮上來，別人卻都沒有，偷錢的不是你是誰？”那人俯首認罪，眾人皆心服。

此故事有趣歸有趣，但無懈可擊嗎？先來看底下的例子。

M 君一人獨中樂透彩頭獎，獎金超過二十億，史上最高。現身領取獎金時，錢還沒領到，調查員卻出現了。“你中樂透彩頭獎那期，買了幾張？”調查員問。“1 張。”M 君回答。“買 1 張就中頭獎？有作弊嫌疑！”調查員口氣不善。“怎麼作弊啊？運氣好不行嗎？”科學辦案，調查員多少懂些統計，他當場展示一假設檢定給 M 君看。

49 取 6 的彩券，中頭獎要 6 碼全吻合，不計順序，因此每張彩券中頭獎的機率為

$$1/C(49,6)=1/13,983,816，$$

將近 1 千 4 百萬分之 1 的機率，的確很難中。現令

$$H_0\text{：M 君沒作弊，}H_a\text{：M 君作弊，}$$

分別表虛無假設及對立假設。“作弊是什麼意思？”M 君謹慎地問。“買通樂透彩公司的員工、會算牌、有預測能力，或任何形式的作假都算。”調查員回答。“那拒絕域能取成什麼？”M 君有點想不透。“先不必管拒絕域是什麼，但任一合理的拒絕域，都該包含你(M 君)中頭獎之事件。”調查員再度說明。“顯著水準 α 取成多少？”M 君再問。“等下再說。”調查員似乎不覺得那很重要。禍從天降，雖並不滿

意調查員的回應，但由於虛無假設是被保護的，擅長統計的 M 君，相信經此檢定，反而可證實他的清白，讓他順利領到獎金，也就不計較了。

在 H_0 為真下，觀測到 M 君中頭獎，由於沒有比這更極端的事件了，故此機率即

$$p\text{-值}=1/13{,}983{,}816。$$

實務上 α 極少取成這麼小，所以在任一合理的 α 下，皆該拒絕 H_0，而接受 H_a，也就是接受 M 君作弊。“你沒作弊誰作弊？”調查員正色道。怎會這樣？M 君愣住了。

僅憑某人身上的錢有油，就認定他偷了賣油條小孩的錢。這種典型包公斷案的手法，雖是千古美談，但其實令人提心吊膽。如果機靈的偷錢者，根本早就躲遠了，而圍觀者中，有人稍早買油條時沒零錢，則找回來的銅錢，不就沾了油嗎？包公因而認定他偷錢，豈不誤判？再仔細想想，在“包公審錢案”裡，除了沒有給出在未偷賣油條小孩的錢之下，身上有沾了油的錢之機率外，包公判定的依據，與調查員由某人中頭獎，便檢定出他作弊，本質上是一樣的。推斷都是憑一假設檢定，包公的有爭議，那調查員的呢？也一樣。若依調查員的論點，則任一位中頭獎者，皆會被“證實”作弊！這樣的假設檢定，豈會令人心服？這麼說，假設檢定不可靠嗎？

首先，因投擲銅板，所獲正面數的多寡，只受銅板正面出現機率的影響，故由投擲後出現的正面數，來檢定正面出現之機率，並無不妥，只要各次投擲，的確相互獨立。但身上擁有沾了油的錢，並不見得是來自賣油條的小孩。就算是，也不見得是偷來的。其次，投擲銅板 20 次，在出現 20 個正面下，懷疑銅板非公正，算是合理。因當銅板為公正，此機率才

$$(1/2)^{20}=1/1,048,576，$$

約百萬分之 1 的機率，相當小。但小機率若遇到大樣本，譬如說有 2 百萬人相繼做此實驗，則其中有人擲出 20 個正面，乃很平常，一點都不該訝異。這點應不難理解。同理，在彩券銷售量夠大的情況下，有人中頭獎，便幾乎是必然，不該連想到作弊。還有一點要注意，對於銅板，誤判其公正性較無妨。但若涉及到人或其他重要的事務，導致犯錯的後果影響較大時，則就須儘量謹慎。總之，運用假設檢定來判定時，萬不可因見到 p-值很小，就理直氣壯地認為結論已定，不必再多說了。

在前述檢定樂透彩是否作弊之例子中，除給出

$$p\text{-值}=P(\text{M 君中頭獎}|\text{M 君沒作弊})$$

外，站在 M 君的立場，應也檢視

$$P(\text{M 君沒作弊}|\text{M 君中頭獎})$$

之大小。否則 M 君將以為調查員故入人罪，辦案完全一廂情

願，這樣他是不會服氣的。要知即使 p-值很小，上述條件機率並不必然也很小。而此條件機率，才是從中頭獎者的角度，與是否作弊密切相關的一個機率。底下來討論此機率值。

處理條件機率，可利用貝氏定理(Bayes' Theorem)。即對二事件 A，B，只要 $P(A)$及 $P(B)$皆不為 0，便有

(1) $$P(B|A)= P(A\cap B)/P(A)=P(A|B)P(B)/P(A)$$

$$=P(A|B)P(B)/(P(A|B)P(B)+P(A|B^c)P(B^c))，$$

其中 B^c 表 B 之餘集。現令事件 A 表 M 君中頭獎，B 表 M 君沒作弊。則事件 B^c 表 M 君作弊。我們想求

$$P(B|A)=P(\text{M 君沒作弊}|\text{M 君中頭獎})。$$

切記 $P(B|A)$與 $P(A|B)$，此二條件機率是不一樣的，不可混淆。為了適用更一般的情況，令

$$p=P(A|B)=P(\text{M 君中頭獎}|\text{M 君沒作弊})，$$

又令

$$q=P(B^c)=P(\text{M 君作弊})，$$

則

$$P(B)=P(\text{M 君沒作弊})=1-q，$$

我們再令

$$P(A|B^c)=P(\text{M 君中頭獎}|\text{M 君作弊})=1，$$

上述條件機率取為 1，乃因我們假設作弊便一定成功。因若作弊不成功，則不會中頭獎，也就沒有後續的檢定問題。當然，若要更一般，可允許 $P(A|B^c)$不為 1。在以上的假設下，由(1)式，即得

(2) $$P(B|A)=p(1-q)/(p(1-q)+q)。$$

回到我們的情況，p=1/13,983,816，但 q 是多少可就未知了。假設 $q=10^{-1}$，代入(2)式，即得

$$P(B|A)=9/13,983,825\approx 0.000000643。$$

假設 $q=10^{-2}$，則

$$P(B|A)=99/13,983,915\approx 0.000007079。$$

假設 $q=10^{-3}$，則

$$P(\mathrm{B}|\mathrm{A})=999/13,984,815\approx 0.00007143。$$

假設 $q=10^{-4}$，則

$$P(B|A)=9,999/13,993,815\approx 0.0007145。$$

假設 $q=10^{-5}$，則

$$P(B|A)=99,999/14,083,815\approx 0.007100。$$

假設 $q=10^{-6}$，則

$$P(B|A)=999,999/14,983,815\approx 0.06674。$$

假設 $q=10^{-7}$，則

$$P(B|A)=9,999,999/23,983,815\approx 0.4169。$$

假設 $q=10^{-8}$，則

$$P(B|A)=99,999,999/113,983,815\approx 0.8773。$$

最後，假設 $q=0$，即 M 君絕對不會作弊，則

$$P(B|A)=1。$$

在 $q=10^{-1}$、…、10^{-7}、10^{-8}，及 0 之假設下，我們分別求出

$$P(B|A)=P(\text{M 君沒作弊}|\text{M 君中頭獎})。$$

這些機率，都比 p-值大很多。且可看出，若 M 君作弊的機率 q 愈小，則在 M 君中頭獎下，並沒作弊的機率便愈大，亦即有作弊的機率將愈小。若 q 與 p(即 M 君沒作弊下，會中頭獎的機率)相比不大，如 $q=10^{-7}$ 或 10^{-8}，則 P(M 君沒作弊|M 君中頭獎)便不算小，此時豈有必要去懷疑 M 君作弊？所以只因

$$p\text{-值}=P(\text{M 君中頭獎}|\text{M 君沒作弊})$$

很小，就見獵心喜，一口咬定 M 君中頭獎是因作弊，則犯錯機率，便絕非以為的 p-值那麼小。

你可能會好奇，如何得知 q 究竟多少呢？對事件的推論，可藉助先驗機率(prior probability，有時只簡單地稱為 prior)，這是貝氏學派(Bayesian)的想法。可由過去的經驗(如

樂透彩曾發生的作弊次數)，及 M 君的資料(職業、生活狀況，及過去行為的紀錄等)等，來推估先驗機率 q。就算不了解 M 君，也可由樂透彩發行單位所掌握的作弊資訊，由全體 q 的估計值，做為 M 君個人的 q 之估計值。一般而言，不論全體或個人的 q，都應極小。因現場隨機開獎，豈那麼容易作弊？對 q 的推估會不會很主觀？當然不無可能。但主觀機率本來就是幾種主要對機率的解釋之一。無論如何，就算不是很精準，如估計 q 介於 10^{-7} 至 10^{-5} 間，而得 P(M 君沒作弊|M 君中頭獎)介於 0.007100 至 0.4169 間，仍比對 q 毫無想法下，能得到更有效的推論。

最後，我們來看一著名的“檢察官的謬誤”(prosecutor’s fallacy)。此為一過度依賴假設檢定，而犯下難以挽回的大錯之實例。

莎莉克拉克(Sally Clark，1964−2007)是家中的獨生女，她父親是位資深警官，母親是位美容師，她與先生同為律師。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。不料這個健康的男嬰，卻在當年 12 月，11 週大時在家中猝死。從悲傷中復原後，於 1997 年 11 月，莎莉又生了一個兒子。豈料 8 週後，1998 年 1 月，不幸的事再度降臨，嬰兒在家中猝死。但不幸還沒結束，因兩次事故發生，都只有莎莉一人在家，她被以殺嬰的罪名起訴。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，乃極不尋常。為了說服陪審團，這絕非猝死，檢察官找來梅鐸(Sir Roy Meadow，1933−)作證。梅鐸是位夙負盛名的

小兒科醫生，且上法庭作證的經驗豐富。

梅鐸向陪審團說明，一家有兩個嬰兒接連猝死的機率，僅有 7,300 萬分之 1，那是多麼微乎其微。只是梅鐸誤解機率，他提供的數據，完全不可信，細節可參考黃文璋(2016a)一文。但陪審團卻接受了梅鐸的證詞。1999 年，莎莉被判無期徒刑，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，最高法院才改判莎莉無罪。只是清白來的太遲，出獄後，莎莉一直處於精神不佳的狀態。2007 年 3 月，她因酒精中毒，死於家中。

我們來重新檢視。如前，令

$$a=P(\text{二嬰兒猝死}|\text{莎莉沒殺二嬰})\text{，}$$

$$b=P(\text{莎莉殺二嬰})\text{。}$$

則仿上述樂透彩作弊之檢定一例的推導，即得

(3) $P(\text{莎莉沒殺二嬰}|\text{二嬰兒猝死})=a(1-b)/(a(1-b)+b)$，

此處依舊合理地假設 $P(\text{二嬰兒猝死}|\text{莎莉殺二嬰})=1$。

梅鐸的機率值雖不可靠，但我們還是先來看，就算採用他所宣稱的機率，莎莉的嫌疑，是否真有那麼大？即取 a=1/7,300 萬。至於 b，以莎莉的背景，取 b=1/1 百萬，都可能太大了。將此二 a，b 代入(3)式，得

$P(\text{莎莉沒殺二嬰}|\text{二嬰兒猝死})=999{,}999/73{,}999{,}999\approx 0.01351$。

此機率可遠比梅鐸提供的 a=1/7,300 萬，著實大多了。更不要說若 b=1/1 千萬，或 b=1/1 億了。

在黃文璋(2016a)一文中，我們提到，莎莉獲釋後，2004 年，有研究指出，a 應介於

1/338,000 至 1/169,000 間，

為梅鐸以為的 a 之 215 倍以上。對此新的 a，仍取 b=1/1 百萬，則得 P(莎莉沒殺二嬰|二嬰兒猝死)介於

999,999/1,337,999 至 999,999/1,168,999 間。

即約介於

0.747383966 與 0.855431869 間。

這樣的機率與小到誇張的 1/7,300 萬相比，大小有如天壤之別。換句話說，在二嬰兒猝死下，莎莉沒殺二嬰的機率，絕非梅鐸所說的那麼小。如此一來，陪審團的決定，可能將大不相同了。

執行假設檢定時，會產生出乎意外的誤差，並不僅限上述只從一個角度看的情況。我們再舉一情況。有位老師，要班上學生各自以投擲，來檢定一銅板出現正面的機率是否大於 1/2，且給出 p-值。以 p 表銅板出現正面的機率。顯然取

H_0：p=1/2，H_a：p>1/2。

K 生與 L 生投擲後，皆得 19 個正面及 1 個反面，但兩人的

p-值卻不一樣。這很奇怪嗎？

K 生事先設訂投擲 20 次，所以出現的正面數有二項分佈 $B(20,p)$。而比出現 19 個正面，至少同樣極端的事件，為出現 19 或 20 個正面。又在 H_0 之下，p=1/2。故 K 生之

$$p\text{-值}=P(\text{投擲 20 次，出現 19 或 20 個正面}|p=1/2)$$

$$=(C(20,19)+C(20,20))/2^{20}=21/2^{20}\text{。}$$

至於 L 生，他乃設定持續投擲，直至出現第 1 個反面，便立即停止。故出現的正面數有負二項分佈(negative binomial distribution) $NB(1,p)$。此時比出現 19 個正面，至少同樣極端的事件，為出現 19，20，…個正面。在 H_0 之下，仍有 p=1/2。故 L 生之

$$p\text{-值}=P(\text{在出現第 1 個要反面前，至少出現 19 個正面}|p=1/2)$$

$$=1/2^{20}+1/2^{21}+1/2^{22}+\cdots=1/2^{19}\text{。}$$

K 生的 p-值，為 L 生的 p-值之 10.5 倍。

此例告訴我們，雖得到相同的正面數與反面數，還要知道是如何投擲的，才有完整的資訊。若只看到 K 生與 L 生所紀錄歷次投擲的結果，正、正、…、正、反完全相同，都是連續 19 個正再 1 個反，便以為數據相同，p-值也必相同，則將可能大錯特錯。要知相同的數據，有些會導致拒絕 H_0，有些卻會導致接受 H_0。因此檢定前得先確定，這些數據究竟是怎麼來的。來源若有異，結論將有能完全相反，不可不慎。

欲對假設檢定進一步了解者，可參考黃文璋(2005)一文。黃文璋(2016b)則為較淺顯的一文，亦可參考。

9 相關性

戰國時代“和氏之璧”的故事，相信聽過的人不少。若想知道詳細內容，當然是上網 google 一下。在某網站對此故事的介紹裡，還列出“相關事件：藺相如完壁歸趙”。喔！這的確相關，也就順便看一下。有些購物網站，當你在查詢某物品時，會出現一句“瀏覽此商品的人，也瀏覽…”，或“買此商品的人，也買了…”，底下則列出若干物品，供你參考。而網站所建議的物品中，有些的確讓你產生興趣，覺得與你所查詢的很相關，有些則令人啼笑皆非，覺得毫不相干，只好過濾掉。如在某“基督山恩仇記”DVD 的銷售網站上，列出五件“買此商品的人，也買了…”，其中兩件是書，一件是 DVD，都還相關，另兩件卻是食品，有鹹蛋黃餅及藍莓乾。無論如何，網路時代，的確讓相關資料的搜尋，更為便捷。生活裡，人們常會主動或被動，接觸到相關事件。如警方辦案，須從千頭萬緒中，找出與案情相關的線索。

在“數學素養研究計畫”中，所訂出的四大數學知識素養領域裡，第一個是“變化與關係”，可見其重要性。變化為大家所熟知。四時行焉，百物生焉，變化不斷。除了物理變化，及化學變化外，還有友情、天氣、經濟等，幾乎什麼都可有變化。至於關係，也是很多事物間，都可拉上關係。

有些是一看便知有關連。如父子關係，身高與體重的關係等。有些關連則不是那麼明顯，但被探索出來。如之前提過的，獅子、老虎、豹，及貓等動物，除了都是哺乳動物外，看似沒什麼特別的關係。但其實牠們在動物的分類裡，皆屬貓科，有共同的祖先，這就不見得普遍為人所知了。又如中學數學裡，方程式的根與係數之關係，及直角三角形三邊長之關係(即畢氏定理)等，都是被探索出來的性質。另外，對多變的事物，人們常企圖掌握其變化。像是對兩個量，若能知道其間的關係，將有助於了解，一量如何隨另一量而變。

有兩種關係很令人感興趣。其一為函數關係，其二為因果關係。函數不過就是一種對應關係。假設有二集合 A 與 B，分別稱為定義域及對應域，若每一 A 中的元素，都恰有一 B 中的元素與之對應，便稱此為一由 A 至 B 的函數。函數關係處處可見，如在台灣，每一國民皆有一身分證字號。由國民對應到身分證字號，便產生一函數。兩個量若有函數關係，其關係便很強，給定定義域中一 x，其對應的 y 便完全確定。至於因果關係，則是指在某假設(或說前提)下，會導致某結論必然成立。所以若不想要該結論發生，就要避免前提發生。例如，酒醉開車易肇事。肇事害人害己，必須避免，於是酒喝到一個標準後，就被禁止開車。有些因果關係較明顯，如任一偶數(前提)的平方，仍為偶數(結論)；有些則須花些功夫去證明，如任給一三角形(前提)，其垂心、外心及重心三點共線(結論，此線稱為歐拉線)。

如果兒子的身高(y)，是其父親身高(x)的一個函數，則父

親有多高，兒子一旦出生，其成年後的身高，便定下來了。只是既然造出這個隨機世界，就不會讓事情這麼單調。由於遺傳，父子的身高的確有相當強的關係，但如果就是函數關係，便失去隨機性了。那隨機世界裡，有沒有因果關係？

2017 年 3 月 19 日，中時電子報有一則新聞的標題是“研究：多養 1 個小孩多活 1 年”。有這種事？來看一下內容。這是由瑞典科學家領導的一項大型研究。對 1911 年至 1925 年間出生，共 70 萬名男性，及 72.5 萬名女性，分析其子女數目、教育程度、婚姻狀態，及死亡時間等。研究指出，男性 60 歲沒有子女者，能再存活約 18.4 年；至少有 1 子女者，則能再存活約 20.2 年，多活約 1.8 年。至於女性，60 歲沒有子女者，能再存活約 23.1 年；至少有 1 子女者，能再存活約 24.6 年，多活約 1.5 年。原來只是比較有小孩跟沒小孩兩個族群，且是比 60 歲以後。並無標題顯示的多 1 個小孩(因)，會多活 1 年(果)的意思。事實上，若有子女，則年紀大後，可能較有人照顧及陪伴，如此身心會較健康、生活品質會較佳，是有助長壽。但年輕時若小孩過多，難免較操勞，再加上經濟負擔較重，是否會有損健康，使有些人連 60 歲都活不到？總之，長壽或短命的原因很多，絕非由單一小孩多寡的因素所決定。況且，人之間的差異性很大，定出比較規則，如比平均壽命，或比長壽的比例，再經一嚴謹的假設檢定，是可判定兩個族群的壽命，能否分出長短。但不宜只針對二人比較。因同樣的因素，對不同的人，效果可能完全相反。換句話說，隨機世界裡，即使討論因果關係，與數學中必然成立的因果關係，意義也可能大不相同。

類似的研究很多，經常可看到報導。譬如說，英國倫敦大學發表一研究成果，指出過度依賴 GPS 導航，會使得大腦變遲鈍。但這是否表示開車時，少用導航系統比較好？恐怕未必，至少由此報導看不出來。如果就這麼簡單，幾千年來，人類所發明的各種工具，可能都宜少用了。常用計算機會變笨，該多依賴筆算。連車子都該儘量避免搭乘，最好多靠走路。要知減少某方面的能力，說不定會增加另些方面的能力。總而言之，經由統計，觀測到的現象，不能當做數學定理，以為建立了一因果關係。

網站搜尋功能日益提升，可快速地找出較相關之事物，供決策者參考。譬如說，Google 網站，由短期間內，擁入多人查詢有關流感的病情，及如何用藥等，推斷流感快爆發了。這種利用搜尋流感的資料，與流感流行有相關性來預測，有時比官方做的預測，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。在著名的“尿布與啤酒”事件裡，美國百貨連鎖店 Wal-Mart，由檢視顧客的購物清單，發現在列有尿布或啤酒的帳單裡，二者同時出現的比例不小。但總不至於說，尿布的使用者，較可能愛喝啤酒。但數據分析，只能做到這裡，想了解原因，得另行調查。有些號稱是大數據專家者，認為在數據可大量且快速分析的時代，只須看重相關性，不必花功夫在探索難以捉摸的因果關係上。像尿布與啤酒，就放在一起賣，何須管為什麼？

查詢流感者，或許不見得染病，而有其他目的，如為了寫報告等，但應有很多是由於自己、家人或朋友染病了；尿

布與啤酒同時出現在帳單裡，即使非同一人使用，總是同一家的人在用。對這類情況，若只在意相關性，大致無妨，但並非永遠都如此。曾有一則新聞的標題是“O 型、射手座、已婚男 最易中彩券”。愛買彩券的人不少，但向來求明牌似乎都不見效果。如今發現中獎與血型、星座、婚姻狀態，及性別相關，那是否幾項都符合的，該趕緊去買？因不是說看重相關性就好，不必在乎因果關係？事實上，單看血型一項，就可了解並非所有相關性都值得看重。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，只是吻合而已。這就如若看到報導，新北市歷來開出最多頭獎，則購買彩券時，須特地跑去新北市嗎？大可不必。因新北市的人口，乃全台各縣市中最多的，理應有最多頭獎落在該市。縣市跟中獎相關性雖高，卻對中獎無影響。

相關性高，但其間沒有因果關係的例子處處可見，我們從黃文璋(2016a)一文中，找出幾個來看。不少人擔心罹患阿茲海默症(Alzheimer’s disease)，尤其是婦女，因有研究指出，病患中約有三分之二是女性。但這項研究其實是說，性別與罹患阿茲海默症的相關性很高，不能就解讀成女性較易罹病。要知阿茲海默症，較多發生在年紀較大時，而女性平均壽命比男性長，因此罹病者以女性居多，乃屬合理。不可忽視其中的“干擾因素”(confounding factor)—壽命。另外，統計顯示磨牙的病患裡，學歷高的較多。只是若將“磨牙與學歷的相關性很高”，解讀成“教育程度較高的人，比較容易磨牙”，就不對了。事實上，學歷較高者，工作職務可能較高，壓力遂也可能較大，有些便以磨牙來舒緩壓力。所以，

是壓力大造成易磨牙，高學歷是無辜的。另一可能性是，教育水準較高的人，可能也較了解健康之重要，當發現自己不對勁地磨牙時，懂得該去看醫生。壓力及知道照顧自己，都可能是干擾因素。還有一類常見的例子。當可樂銷量大時，醫院腸胃科的門診人數也較多。看到此相關性高的統計結果，並不必因此就認為喝可樂易傷腸胃。因可樂銷量大，往往是天氣炎熱時。天氣炎熱，使食物易腐敗，導致易吃壞肚子。氣溫為一干擾因素。

總之，在不少情況下，絕不可只看到相關性高，就以為有什麼大發現。因果關係並非都可不去追究，知道相關性就足矣。但這也不表相關性高這一資訊沒用。可由相關性高的因素出發，進一步探索其中是否有由因果關係。警方辦案，當然會從較相關的人、事、時、地、物開始查，這是合理的，具統計裡“最大概似法”(method of maximum likelihood)的想法。但不能僅停留在相關性較高的因素，而棄因果關係於不顧，這樣就常會進入死胡同。

到此你可能好奇，什麼是相關？這其實並無嚴格的定義。同姓是相關，所謂五百年前是一家；同鄉、同事、同學、同行等，都屬相關。二事物在網路上同時出現是相關，經常同時出現，會被認為相關性高。但當然也不一定。同學多年，所以名字常連在一起，看起來相關性很高。不過可能就僅止於這樣而已，彼此一點都不熟，毫不相干。

統計裡引進相關係數(correlation coefficient，有時只稱correlation)，以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(uncorrelated，或稱 no correlation)，為一相當實用的“度量器”。

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)所首創，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，則便有

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

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

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。兩隨機變數 X 與 Y 的相關係數要存在，先決條件是，X 與 Y 之

期望值皆存在，及變異數都存在且都不為 0。變異數會為 0，就是常數了。如果要較嚴密地講，可說此時隨機變數為一常數的機率為 1，但並沒太大必要。

可看出 Cov(X,X)=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$(X-\mu_X)(Y-\mu_Y)$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$(X-\mu_X)(Y-\mu_Y)$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。事實上，當 ρ=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 ρ=1，則存在常數 a，b，其中 a>0，使得 $Y=aX+b$；若 ρ=-1，則存在常數 a，b，其中 a<0，使得 $Y=aX+b$。

另外，當 X 與 Y 獨立時，$\mathrm{Cov}(X,Y)=0$，因而 $\rho(X,\mathrm{Y})=0$。但其逆不真。也就是說，共變異數(或相關係數)為 0，此時稱兩隨機變數為無相關(uncorrelated)，並不導致二變數必獨立。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，知道 X 後，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，如黃文璋(2007b)一文之例 4 亦為一例。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

$$r=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}\text{ 。}$$

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

$$r = 16/(20 \cdot 80)^{1/2}=0.4，$$

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，說不定還負的。至於婦女的教育程度與小孩數，也有可能是負相關。

10 隨機性

我們常提到“隨機”，如隨機現象及隨機變數等。其中的隨機一詞，定義是很寬鬆的。簡言之，只要不是一成不變即可。假設一銅板有正及反兩個面，往上一拋，觀測落地後朝上的為正面或反面，此便是一隨機現象，因事先不確定那一面會朝上。若銅板兩個面皆為正，則投擲後，朝上的必然是正面，沒有其他可能，所以這並非隨機現象。但亦有人稱

此為一退化的隨機現象。即只要視一成不變為退化，則所有的現象，便無不為隨機現象。在隨機現象裡，用來投擲的銅板或骰子等，都不一定要公正，但仍要符合該有之規定，如各面出現的機率和為 1。

不過對於統計檢定或民調裡，所提到之隨機取樣，其中的隨機，往往有特別的涵義。底下為幾種常見的情況。第一種是隨機樣本。例如，為檢定某廠牌電池的壽命，首先便是隨機取 n 個來測試，因而得到 X_1，⋯，X_n，其中 X_i 表第 i 個受測電池的壽命，i=1，⋯，n。一般會假設電池品質一致，且各電池的測試互不干擾，亦即設 X_1，⋯，X_n，有相同的分佈，且相互獨立。則 X_1，⋯，X_n，便稱為一組隨機樣本。第二種情況是簡單隨機抽樣(simple random sampling)，乃針對有限的母體。欲取一組樣本數為 n(n 當然要小於母體數)之樣本，若母體中任一組可能的樣本，皆有相同的機率被抽到，便稱此抽樣步驟為簡單隨機抽樣，而獲得的樣本，則稱為簡單隨機樣本(simple random sample)。須注意的是，是要求每一組樣本數為 n 之樣本，皆有相同的機率被抽到，而不僅是每一個樣本，皆有相同的機率被抽中，前者的條件較強。即若母體數為 m，$m>n$，則總共的 C(m,n)組可能之樣本，每一組被取中的機率，皆須為 1/C(m,n)，如此所得才是隨機樣本。如果自母體中，每次依序隨機取一樣本，取出後皆不放回，即第 1 次，母體中每一樣本被取中的機率為 1/m，第 2 次每一樣本被取中的機率為 1/(m-1)，以此類推，則 n 次後，所得便是一組樣本數為 n 之簡單隨機樣本。這是一較易執行之實際取樣方式。

如何檢定一銅板是否公正？當然就是先隨機投擲。此處的隨機，就是指將獲得一組隨機樣本。更明確地說，投擲同一銅板(其涵義為各次得到之結果，有相同的分佈)n 次，各次投擲假設互不干擾(其涵義為相互獨立)。有趣的是，在隨機投擲下，其中又會含有不少隨機性。例如，假設隨機投擲一銅板(並不須是公正)4 次，得到 2 個正面。此 2 正面會出現在那裡？C(4,2)=6，從 4 次中取 2 次來放正面，6 種可能的位置，出現之機率會相同，即皆為 1/6。一般若一銅板隨機投擲 n 次後，得到 k 個正面，此 k 個正面如何散佈？將有如母體數為 n，從中採簡單隨機抽樣，以得一組 k 個樣本。即總共的 $C(n,k)$組可能之樣本，每一組被取中的機率，皆為 $1/C(n,k)$。

A 君想讓 B 君相信他手上的銅板為公正。隨機投擲 20 次，假設出現 10 個正面，顯然可通過銅板是公正之檢定。只是 B 君注意到，前 10 次投擲出現 7 個正面，後 10 次投擲卻僅出現 3 個正面。這麼偏頗？銅板真的公正嗎？B 君懷疑。A 君不計較，重新投擲銅板 20 次，這回剛好出現 1 正 1 反，…，1 正 1 反。仍出現 10 個正面，且前後 10 次各 5 個，這樣總沒話講了吧！A 君心想。不料細心的 B 君再度指出，10 個正面，皆出現在奇數次投擲，偶數次投擲，則 1 個正面也無。B 君的一再質疑，反使 A 君想通了。他對 B 君說，10 個正面，就是會出現在某 10 次投擲，你豈非永遠可質疑，為何僅這 10 次出現正面？B 君總算不說了。

事實上，10 個正面集中在前 10 次，或奇數次投擲，這些是較明顯的，很難不被注意到。有如 49 取 6 的樂透彩，

若頭獎號碼開出 1，2，3，4，5，6，可能會很轟動，傳誦多年。但其實這組號碼，與任何一組，如看起來很普通的 9，14，23，29，35，46，及 4，10，21，23，38，46 都一樣，出現的機率，皆為 1/C(49,6)=1/13,983,816，只是 1，2，3，4，5，6 這組，較引人注目而已。

投擲銅板 20 次，在出現 10 個正面下，10 個正面皆出現在奇數次(或前 10 次)投擲，其機率為 1/C(20,10)=1/184,756，的確相當小。所以若出現此狀況，有人感到懷疑，並非完全不合理。大膽懷疑無妨，惟須小心求證。要知投擲銅板，若出現的正面數過於偏頗，會讓人認為銅板非公正；但若出現過於巧合的結果，也是會讓人懷疑有作假。現考慮投擲次數較一般的情況。假設投擲一銅板 $2n$ 次，且出現 n 個正面。這本來還好，但仔細一看，n 個正面，都發生在奇數次投擲，其機率為 $1/C(2n,n)$。當 $n=1$，2，3，此機率分別為 1/2，1/6，1/20，愈來愈小，且隨著當 n 之增大，將漸減至 0。換句話說，n 不大時還好，n 愈大恐會愈起疑。1 正 1 反交錯，看起來很“公平”，卻不但不會讓人相信銅板確實公正，反而會懷疑是否真做到隨機投擲？底下來看一些例子。仍先看公正銅板。

隨機投擲一公正銅板 20 次，給定出現 10 個正面。令

$M_i=P$(前、後 10 次投擲中，各出現 i 及$(10-i)$個正面)，

$i=0$，1，…，10。

則

M_i=C(10,i)C(10,10-i)/C(20,10)

=C(10,i)^2/C(20,10)，i=0，1，…，10。

由於 M_i=M_{10-i}，因此只需給出 M_5，…，M_{10}，即可。之前已給

C(20,10)=184,756，

再求出

C(10,5)=252，C(10,6)=210，C(10,7)=120，

C(10,8)=45，C(10,9)=10，C(10,10)=1。

即得

M_5=252^2/184,756=63,504/184,756≈0.344，

M_6=210^2/184,756=44,100/184,756≈0.239，

M_7=120^2/184,756=14,400/184,756≈0.078，

M_8=45^2/184,756=2,025/184,756≈0.0110，

M_9=10^2/184,756=100/184,756≈0.000542，

M_{10}=1/184,756≈0.00000542。

由上述 M_i，可看出前、後 10 次投擲中，各出現 5 個正面，看起來是很均衡，但發生機率約為 0.344，只比 1/3 稍大些。出現不均衡的可能性反而大多了，約 0.656，接近 2/3。而前、

後 10 次投擲中，有 1 出現 6 個正面，有 1 出現 4 個正面，發生機率約為 $2M_6 \approx 0.478$。原來有點偏差，才是最可能的。但並不致於過度偏差。如 10 個正面，都集中在前、後的某 10 次投擲中，這麼極端的情況，其發生機率僅約 $2M_{10} \approx 0.0000108$，根本很少會出現。

我們再令

$N_i = P$(前、後 10 次投擲中，有一出現至少 i 個正面)，
$i=0$，1，…，10。

則 $N_0=N_1=\cdots=N_5=1$，這是當然的。又

$$N_i=2\times(M_i^2+\cdots+M_{10}^2)/C(20,10)，i=6，\cdots，10。$$

故

$N_6=121{,}252/184{,}756 \approx 0.656$，

$N_7=33{,}052/184{,}756 \approx 0.179$，

$N_8=4{,}252/184{,}756 \approx 0.0230$，

$N_9=202/184{,}576 \approx 0.00109$，

$N_{10}=2/184{,}576 \approx 0.00001$。

投擲銅板 20 次，總共出現 10 個正面，銅板的公正性，看起來並沒問題。但在給定出現 10 個正面下，若前、後 10 次投擲中，有一出現過多(如 8、9 或 10 個)的正面，是可懷疑投擲過程，並非如所宣稱的隨機。至於若要釐清，就去執行一

假設檢定，以進一步確認。

在上例中，N_7不算小，有 0.179，N_8雖較小，卻也仍有 0.0230。顯示投擲一公正銅板 20 次，在給定出現 10 個正面下，前、後 10 次投擲中，出現的正面數為 7 及 3，3 及 7，8 及 2，或 2 及 8，偏離期望值 5 雖遠些，但發生的可能性都尚不算很小，主要是投擲數 20 不是很大。假設投擲次數 40，且依然出現半數 20 個正面。則可求出對應的

$$N_{15}=2\times(C(20,15)^2+C(20,16)^2+C(20,17)^2+C(20,18)^2$$

$$+C(20,19)^2+C(20,20)^2)/C(40,20)\approx 0.00192。$$

此時前、後 20 次投擲中，有一出現至少 15(75%)個正面之機率，便大幅下降。

最後，來看一非公正銅板之例。假設投擲一出現正面機率為 0.88125 之銅板 56 次，得到 50 個正面。由於期望所得之正面數為 49.35，所以對所宣稱出現正面機率為 0.88125，並沒什麼好挑剔的。令 M_i及 N_i的定義如上，但此處將 56 次投擲，分成前 28 次，及後 28 次兩部分。則

$M_{25}=C(28,25)^2/C(56,50)=10,732,176/32,468,436\approx 0.331$，

$M_{26}=C(28,26)C(28,24)/C(56,50)=7,739,550/32,468,436\approx 0.238$，

$M_{27}=C(28,27)C(28,23)/C(56,50)=2,751,840/32,468,436$

≈ 0.0848，

M_{28}=C(28,28)C(28,22)/C(56,50)=376,740/32,468,436≈0.0116。

由此得 N_{25}=1，且

N_{26}=21,736,260/32,468,436/32,468,436≈0.669，

N_{27}=6,257,160/32,468,436≈0.193，

N_{28}=753,480/32,468,436≈0.0232。

跟之前投擲銅板 20 次出現 10 個正面之例比較，如今前、後 28 次投擲中，各出現 25 個正面，此最均衡的狀況，發生之機率 M_{25} 最大，約為 0.331；至於愈偏差的機率則愈小。又前、後 28 次投擲中，除非有一全出現正面(即有 28 個)，此機率約為 0.0232，否則都不算太偏頗。

11 結語

處在此隨機世界，統計是做決策之一不可缺的依據。統計的重要性，已不必贅言。統計很難嗎？在黃文璋(1998)一文裡，我們以數種指標來說明，華人在世界統計領域的表現極優異。底下為指標之一。統計界有一個大獎—考普斯會長獎(COPSS Presidents' Award)，由幾個重要的國際統計組織所支持。COPSS 之全名為 The Committee of Presidents of Statistical Society)，每年頒給一位在統計方面有重要貢獻之 40 歲以下的學者，有人認為此獎相當於統計界的諾貝爾獎。自 1981 起頒發，至 2016 年，共有 36 位獲獎者。其中華人

有 9 位，佔了 4 分之 1，比例高的驚人。華人不乏數學訓練佳，根基紮實者，加上一些其他的特質，使不論在統計的學術或實務上，屢有表現傑出的華人。但這可不表擅長數學者，就能輕易駕馭統計。我們強調過了，統計學不等同於數學，統計學絕非數學的子集合。沒錯，統計的確常會用到數學，是得仰賴“數據”。但若不了解隨機性，以為統計不過是較簡單的數學，而數據就是數字而已，則自以為運用統計所得出的結論，若令人存疑，並不足為奇。統計素養若不足，也難以精準地掌握統計相關文獻之內涵。

由於生活中離不開統計，因此媒體上，涉及統計的報導，經常可見。底下是一些隨手找到，1 個多月內，幾則與統計有關之新聞的標題：

1. 研究：一天一杯茶 有助預防失智症(2017 年 3 月 20 日中央社)。

2. 太沉重！大數據：失智患者平均餘命長(2017 年 3 月 21 日聯合報)。

3. 太瘦會死！研究：老人 BMI 過輕 死亡率飆高 5 成(2017 年 4 月 7 日中國時報)。

4. 疑遭仙人跳 警察大數據資料無所遁形(2017 年 4 月 27 日中央社)。

5. NBA／詹姆斯賽前發言被噓 騎士仍打爆綠衫軍(2017 年 4 月 6 日聯合報)。

6. NBA／數據不說謊 火箭貝佛利嗆衛少濫投才高分(2017 年 4 月 26 日聯合報)。

第一則新聞“一天一杯茶 有助預防失智症”，內容如下：

> 根據新加坡國立大學的最新研究，1 天 1 杯茶有助於降低罹患失智症的風險。…。這項發表在“營養、健康和老化”期刊(Journal of Nutrition，Health and Ageing)的報告指出，每天一杯茶，有助降低年長者認知損害風險 50%。對擁有較易罹患阿茲海默症基因的人，1 天 1 杯茶，更可以減少失智症風險達 86%。…。針對 957 名年齡 55 歲或以上的華人，進行為期 2 年的研究。…。失智症的藥物治療遙不可及，也還沒有令人滿意的結果。這項研究僅限於紅茶、綠茶和烏龍茶等茶葉沖泡的茶飲。

研究結果指出，(華人)年長者，僅靠每天一杯茶，兩年後就能降低認知損害風險 50%。兩年降低一半，等比級數，那是否 4 年後降低 75%？6 年後降低 87.5%？8 年降低 93.75%？也就是(華人)55 歲後，只要持續每天一杯茶，便不必擔心認知損害了。但真有這麼神奇？因報導中不是說，失智症的藥物仍未誕生！

第二則新聞“大數據：失智患者平均餘命長”，內容如下：

> 失智病患比一般人更長壽！…。曾任衛生署統計室主任，現任全球醫資與亞洲環宇生技大數據顧問黃旭明指出，以失智者的死亡率推估活年數，發

> 現失智者普遍高齡。…，50 歲失智者的平均餘命達 33 年；60 歲者有 25 年；70 歲者有 17 年；80 歲者有 11 年；90 歲者有 5 年，…。黃旭明說，失智症患者比一般人更長壽，對照護者而言，應該及早做好照顧與財務的規畫。…。

失智問題令人重視，因此相關的研究不少。失智後可能無法正常言語、容易迷路、喪失生存動力、喪失長期記憶、難以自理生活，及行為異常等。病情嚴重者，還會逐漸喪失身體機能。疾病的進程固然因人而異，但由於有前述諸多症狀，一般而言，難以存活太久(維基百科上說，確定診斷後的平均餘命是 3 到 9 年，確診後存活超過 14 年者少於 3%)。光是“容易迷路”，就不知會造成多大的危險。但此報導，卻指出失智者，其平均餘命較一般人長，豈不違反常理？附帶一提，近年來，人們以往“數據會說話”那句口頭禪，已逐漸轉為“根據大數據”，或“根據大數據分析”。只是人人琅琅上口的“根據大數據”，雖看起來莫測高深，卻有如說“利用數學”，而到底利用些什麼？其實幾乎沒說。

第三則新聞“太瘦會死！研究：老人 BMI 過輕 死亡率飆高 5 成”，內容如下：

> …。國內最新統計顯示，超過 60 歲的長者，BMI(身體質量指數)若小於 18.5，死亡率較正常體位者“高出 5 成”，主因是過瘦的長者常會“肌少症”上身，提高身體發生意外、慢性發炎、感染和罹癌風險，死亡率相對提高。該項由國衛院進行的

> 最新調查顯示，60 至 69 歲，BMI 小於 18.5 的長者，相較同年齡層、體位正常者，死亡率高出 51%，70 歲以上更高出 60%，反轉過往“過胖才易致死”的觀念。…。統計顯示，國內目前 65 歲以上老人約佔總人口 13.2%，當中肌少症盛行率約 7.9%，等於約有 24 萬名老人罹患肌少症，容易因沒力而發生跌倒等意外事件。…。

這是國家衛生研究院(簡稱國衛院)所做的統計，應很有權威才對。但事情並沒講清楚。要知統計裡找到關連性強的因素，不表就找到因果關係。BMI 為體重(單位取為公斤)除以身高(單位取為公尺)的平方。BMI 小於 18.5，對身高 1.7 公尺的人，表體重少於 53.465 公斤；對身高 1.6 公尺的人，則表體重少於 47.36 公斤，均相當輕。只是上了年紀的人，體重過輕，往往是身體有些毛病，導致無法順利吸收食物的養分，遂骨瘦如柴了。真實的情況，應是健康不好，造成死亡率較高。過瘦是健康出問題的果，而不是因。若健康亮紅燈的老人，看了上述報導後，開始大吃，企圖增如體重，恐將讓身體產生更多毛病。

第四則新聞“疑遭仙人跳 警察大數據資料無所遁形”，內容如下：

> 新北市男子疑遭“仙人跳”，被強行擄走拘禁毆打。警方今天表示，透過刑案關聯性大數據資料庫，循線逮捕 1 女 5 男…，函送檢方偵辦。新北市三重警察分局…表示，大數據可作為犯罪研究或調

> 查的基礎，讓原本預測分析的技術有更多發展面向。關聯性分析，不僅可用在刑案資訊上，社群網站交友等，也是偵辦刑案的利器。…。大數據的分析，普遍應用在各領域中。…。利用大數據分析偵辦各類刑案，可提升科技偵查能量，面對龐大的資料量，如能加速找出資料間的關聯性，將資料轉為資訊，就能加速得到線索，成為破案關鍵。

又是大數據！給出本則報導，乃是要再度提醒，二因素的相關性高，不表示其間真有很強的關係。因此所謂“關聯性分析”，運用時要很謹慎。給一例子。選舉到了，那位候選人的當選機率較大？是網路搜尋人氣較高的那位嗎？有可能，但並不必然。2016 年美國總統大選，有媒體依民眾上網搜尋柯林頓(Hillary Clinton，1947−)的次數，比搜尋川普(Donald Trump，1946−)的次數多，預言柯林頓將當選。但亦有媒體在投票前夕，依“為川普祈禱”(Pray for Trump)及“為希拉蕊祈禱”(Pray for Hillary)來搜尋，發現前者的數量，較後者高出許多，因而預言川普將獲勝。結果依第二種搜尋做的預測對了。其實說穿了，對勢均力敵的兩位候選人，就算以投擲銅板來預測，都會有約一半的人成為神算子。

第五則新聞“NBA/詹姆斯賽前發言被噓 騎士仍打爆綠衫軍”，內容如下：

> …，根據 NBA 統計資料顯示，騎士本季只要帶著 15 分以上的優勢進入下半場，他們全數獲勝、戰績是 17 勝 0 敗。…。

克里夫蘭騎士(Cleveland Cavaliers)，是一支位於美國俄亥俄州克里夫蘭的 NBA 籃球隊。以相對頻率來解釋機率，是一常見的作法。由於過去上半場領先 15 分以上的有 17 場，且全數獲勝，所以，若之後有某一次比賽，上半場領先 15 分以上，是可預測該場贏的機率為 1(=17/17)。當沒有其他資訊時，以此法來預測並非不合理。只是一方面，樣本數並不算大；另一方面，騎士隊上半場領先 15 分以上的那些球賽，也並非真的為獨立且有共同分佈。因而這樣的預測，只能供參考而已。畢竟球賽各場的狀況總有所不同，不像投擲銅板那麼單純。所以，若某次騎士隊上半場領先 15 分以上，結果卻輸球，並無須太訝異，且此也完全不是統計的錯。球賽裡類如上述的報導很多。像是有則“NBA》20 年鐵律 勇士橫掃拓荒者就能奪冠？(2017 年 4 月 26 日中國時報)之新聞，內容如下：

> 根據“ESPN”報導，過去 20 年只要能在季後賽橫掃拓荒者的球隊，最終都能拿下總冠軍，勇士能成為這項鐵律的第 4 支摘冠隊伍嗎？…。

金州勇士(Golden State Warriors)，是一支位於美國加州奧克蘭的 NBA 籃球隊；波特蘭拓荒者(Portland Trail Blazers)，則是一支位於美國奧勒岡州波特蘭的 NBA 籃球隊。NBA 的季後賽，在 1999、2001，及 2002 年，各有一支球隊，於連贏拓荒者 4 場(季後賽採 7 戰 4 勝制)後，最終都得年度總冠軍。記者“根據大數據”，搜尋到此趣聞。但你當然知道，對於這種“鐵律”，千萬不必太當真。事實上，只要有心，可搜

尋到更多某幾支總冠軍球隊，共同具備之的特徵，但都宜僅視為趣聞。

第六則新聞“NBA/數據不說謊 火箭貝佛利嗆衛少濫投才高分”，內容如下：

> …。火箭後衛貝佛利(Patrick Beverley)是系列戰球隊能取勝的功臣之一，今天對上雷霆的第 5 戰，他賽中曾和對方主將衛斯特布魯克(Russell Westbrook)產生爭執，…，賽後他證實“衛少”嗆聲豪取 40 分，自己則回敬那是濫投而來。貝佛利說：“他跟我說他拿了 40 分，說根本沒人守得住他，我回答他類似‘喔！那很好啊，你出手 34 次才換來的。’”…。貝佛利表示，人都會說謊，“但數據不會。”言下之意嗆衛斯特布魯克的得分雖高，但出手次數實在太多。…。貝佛利表現亮眼，拿下 15 分、8 籃板之外，全場投 10 中 6。衛斯特布魯克全場出手 34 次，有 18 次在三分線上出手，是他季後賽最多三分出手的數據，但只進了 5 球。

休士頓火箭(Houston Rockets)，是一支位於美國德州休士頓的 NBA 籃球隊；奧克拉荷馬雷霆(Oklahoma City Thunder)，則是一支位於美國奧克拉荷馬州奧克拉荷馬市的 NBA 籃球隊。雷霆隊的主將衛斯特布魯克個人得分常很高，甚至在籃板及助攻的數據也常很亮麗，屢獲媒體贊揚。但籃球比賽要看團隊表現，僅個人數據好，並不一定能贏球。衛斯特布魯克出手次數極多，顯然隊友過度仰賴他，常將球傳給他。但

他得分雖高，命中率卻不見得高，因而球隊總得分，尚不足以撐高獲勝率。除了上述與火箭隊季後賽的第五戰，他命中率不高外，在 2017 年 4 月 19 日，與火箭隊季後賽的第二戰，衛斯特布魯克拿到驚人的 51 分。但他 43 投 17 中，三分球 11 投 2 中，最後輸球。數據會說話是沒錯，但貝佛利點出了該聽懂數據說的話。要知數據不會說謊，若以為數據說謊，其實應是人在說謊。能聽懂數據說的話之球員，可不只貝佛利。2017 年 4 月 29 日，聯合報有一則標題是“NBA/季後賽喜歡打客場 詹姆斯愛克服逆境”之報導，內容如下：

> …。32 歲的詹姆斯(LeBron James)已在連續 27 個季後賽的系列賽中，至少都贏過 1 場客場比賽，下次對到暴龍將有望延續紀錄。不過，詹姆斯表示，過去的紀錄並不意味著他將在這輪的系列賽，至少贏得一個客場，對他而言，只想努力贏得每一場比賽，試圖幫助球隊贏球。…。

多倫多暴龍(Toronto Raptors)，是一支位於加拿大安大略省多倫多的 NBA 籃球隊。可看出詹姆斯不但球打得好，也能聽懂數據說的話。

統計素養並非一朝一夕可培養出，須靠不斷思索統計裡的各概念，方能逐漸較了解統計的內涵。對統計思維有興趣者，可參考黃文璋(2009)一文。初學統計者，統計概念經常夾纏不清，追根究底，乃因很多時候未弄清楚機率的意義，尤其是條件機率。可參考黃文璋(2000)、(2010)、(2011e)及(2011f)。時至今日，人們面臨的資料，不但愈來愈大量，也

愈來愈複雜，亦即常會遇到大數據。處理大數據，往往不能只靠統計，而得與資訊結合。但不論數據如何大，處理起來總脫離不了統計。具備足夠的統計素養，在大數據時代，更顯必要。

參考文獻

1. 橋本忍(2006a)。複眼的映像—我與黑澤明(張秋明譯)。大家出版社，新北市。
2. 黃文璋(1998)。統計與棒球。中國統計通訊，9(8)：19-28。
3. 黃文璋(2000)。瞻前顧後。統計薪傳，1(1)：83-99。
4. 黃文璋(2005)。統計顯著性。數學傳播季刊，29(4)：29-38。
5. 黃文璋(2006b)。統計裡的信賴。數學傳播季刊，30(4)：48-61。
6. 黃文璋(2007a)。統計裡的估計。數學傳播季刊，31(2)：3-20。
7. 黃文璋(2007b)。統計裡的關係。數學傳播季刊，31(1)：49-67。
8. 黃文璋(2009)。統計思維。數學傳播季刊，33(4)：30-46。
9. 黃文璋(2010)。機率應用不易。數學傳播季刊，34(1)：14-28。
10. 黃文璋(2011a)。統計與考古。科學人，118期(2011年12月號)：33。

11. 黃文璋(2011b)。機率統計考題探討。黃家小館(http://www.stat.nuk.edu.tw/huangwj)

12. 黃文璋(2011c)。關於中央極限定理之圖示。黃家小館(http://www.stat.nuk.edu.tw/huangwj)。

13. 黃文璋(2011d)。庶民中央極限定理。黃家小館(http://www.stat.nuk.edu.tw/huangwj)。

14. 黃文璋(2011e)。認識機率。數學傳播季刊，35(2)：32-44。

15. 黃文璋(2011f)。對機率要有信心。黃家小館(http://www.stat.nuk.edu.tw/huangwj)。

16. 黃文璋(2014a)。PR值與百分位數。黃家小館(http://www.stat.nuk.edu.tw/huangwj)

17. 黃文璋(2014b)。談不確定性。黃家小館(http://www.stat.nuk.edu.tw/huangwj)

18. 黃文璋(2014c)。談信賴區間。翰林數學天地，36期(2014年4月號)：7-10。

19. 黃文璋(2015a)。數據素養。黃家小館(http://www.stat.nuk.edu.tw/huangwj)

20. 黃文璋(2015b)。談中位數。科學人，164期(2015年10月號)：32。

21. 黃文璋(2016a)。談統計誤差—假設檢定篇。黃家小館

(http://www.stat.nuk.edu.tw/huangwj)。
22. 黃文璋(2016b)。機率與統計在高中。中國統計學報 54(2)：43-61。