統計與因果關係

黃文璋

國立高雄大學應用數學系、統計學研究所

1972 至 1974 年，英國曾進行一項醫學研究，二十年後做後續追蹤調查，收集到很多寶貴的數據。有人分析其中婦女抽菸與死亡兩變數，得到抽菸者之死亡率，顯著高於不抽菸者之死亡率的推論。數據會說話，難道抽菸導致較長壽？這當然違反一般人的認知。如果仔細檢視那些數據，將發現其中有一重要的因素不能忽略，即年齡。在初次調查時，高齡層中，抽菸的比例很低。這應很合理，不少菸槍於上了年紀後，才願意相信抽菸有礙健康，於是紛紛戒菸。只是經過二十年，不論抽菸與否，年紀大的總是較難存活。因此有了年齡這個因素後，本來是年紀大的比較多死亡，呈現出來的，卻是死亡者中，較多屬於不抽菸者。想通這點後，將初次調查時，65 歲以上那些人的資料去掉，則抽菸者之死亡率便比較高了。

統計是現代做決策一重要的依據，不過經由統計，常無法判定兩變數間是否有因果關係。諸如上述那類例子很多。像是英國廣播公司曾報導一項台灣學者所做的研究。分析台灣自 1999 至 2008 年，1856 位 65 歲以上獨居老人的生活習性，發現經常逛街較不常逛街者，存活率高 27%。難道這是種對健康有益的“購物療法”？如果真有效，一身是病者，

是否該強忍病痛，多逛街以延年益壽？可能未必。說不定正是因身體較佳的獨居老人，才方便自行逛街。因果關係可能是倒過來。就算逛街真有好處，但經濟狀況不佳者，是否就不易藉此強身？事實上亦有人指出，即使獨居，只要樂於外出和別人打交道，減少孤寂感，就有助健康了，逛街不過是這種方式之一。再給一例。有人分析台灣大學入學新生的來源，發現來自台北市大安區之比率為全國最高。想要小孩進台大的，是否該效法孟母三遷，搬到大安區？這就不見得了。針對此項統計報導，台大校方表示，居住大安區讀台大的比率會較高，可能是因該區公教人員比較多，而他們一向較重視孩子的教育。關鍵是父母較重視小孩的教育，而非住在那裡。

不能判定因果關係，那做統計分析有何用？用途還是不小。曾有美國某大型連鎖超市，探勘顧客購物清單，發現啤酒與尿布同時出現的比例很高，於是將啤酒的陳列位置與尿布放在附近，結果兩者銷售量皆提高。難道是使用尿布者，較愛喝啤酒嗎？顯然不是。真正原因有可能是，若家中有嬰兒，當爸爸的便常得留家幫忙照料，不能老往外跑，為打發時間只好看看球賽轉播之類的節目。於是買尿布時，順便買啤酒，以便看電視時享用。或者反過來，買啤酒時順便買尿布，才不會球賽看一半，被太太差遣去買尿布。經由統計分析，能找出那些變數間相關性較高。警方辦案也常如此，從大量的數據中，找到種種線索。但相關性較高的變數間，是否有因果關係，就須更進一步的探討方能得知，不可驟下結論。今年５月，有一則美國麻省理工學院所做的研究報導：

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。有些離婚的單親媽媽看了可能有些不安，覺得當年做了錯誤的抉擇。但如何彌補？是否該趕快再婚，以脫離單親媽媽的狀態？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？要知孩子的成就，乃與遺傳及成長環境關係密切，因此這兩種作法，不但不見得有用，還有可能讓孩子將來的發展更差。