

統計與預測

黃文璋

國立高雄大學應用數學系

我們常對未來做各種預測。近從預測今年國中基測建中最低錄取分數，遠至預測 100 年後，台灣每年新生兒數目。宇宙的運轉有必然性及隨機性交錯著進行。放開手上的銅板，若忽略空氣阻力，則可算出落地所需時間，這是必然性，但那一面朝上就不得而知，這是隨機性。又如我們可以準確算出下次金星凌日的時間（2012 年 6 月 8 日），但颱風走向，就常預測不準。有些預測方式，不是那麼有科學依據，如依靠擲筊、水晶球等。有些預測則採用現代統計方法：先收集資料，再整理資料，然後分析資料，最後給出推論。對於隨機現象，即使掌握的資料很精確，用的方法很好，預測是不易永遠正確的。例如，假設投擲一銅板 100 次，得到 64 個正面，36 個反面。統計的許多方法包括直觀上，都會導致以 0.64 來估計此銅板出現正面的機率。如果要預測下一次投擲會出現那一面，顯然是要說正面。但你知道，還是可能會判斷錯誤。連這麼簡單的正、反兩面二選一，都無法預測精準，又如何能正確預測明年的經濟成長率呢？這樣說來統計有何用呢？其實不然。統計方法所給的是在已知條件下的最佳決策，對於投擲銅板預測出現那一面的問題，正面仍是最佳選擇。另外，實驗次數較大時，便更能凸顯統計的功能。例如，投擲 100 次，那一面出現較多便獲勝，則更要押正面，獲勝機率大約是 0.9971，平手機率約為 0.0013。但即使如此，還是有約 0.0015 的機率，反面數可能較多。

有些隨機量，二者間是有關連的。如知道父親身高想預測兒子身高。此時可由所收集到的父親(X)、兒子(Y)身高的數據，建立一 X 與 Y 間的機率模型。比較合理的模式，給定 X 之值後，Y 應非一定值，而是有某一機率分佈。但往往為了簡便，以某一與 Y 有關的量如期望值，來當做 Y 之預測值。另外，一個機率模型用久之後，如果懷疑是否仍有錯，可能便要做一統計檢定了。

由於預測不可能百分之百精準，因此給出預測值外，通常也要給出誤差大小。例如，民調結果的報告，除了要說明樣本如何產生等取樣過程，其中還必加上底下這類附註：在百分之九十五的信心水準下，抽樣誤差在正負三點二個百分點以內。意思是說，實際值與所提出的預測值，誤差在 3.2% 內的機率為 0.95。對於隨機現象，並無法保證預測的誤差有多大，只能給出誤差不超過某一值之機率。

有時想要預測，有時又不想被人預測中。如提款卡密碼，樂透彩中獎號碼等。隨機產生的號碼，是最難被猜中的。你選 1,2,3,4,5,6 為密碼，以為沒人想得到，但偏偏人同此心，不少人一試便是這組。樂透彩不論開出多少期，下一期每組號碼出現機率仍均相同。從 n 個數字中產生 r 個不重複且順序不計的號碼，依目前樂透彩隨機產生中獎號碼的方式，資訊理論指出，會有最大的不確定性，亦即是最難預測中的。