評比宜有統計概念

黃文璋

國立高雄大學統計學研究所

外地人要到高雄旅遊，有什麼好吃的？先查一下高雄最佳美食餐廳榜。人們常在評比，如台灣十大經典魅力漁港的選拔，世界大學的排名等。只是有時排名很前的，根本名不見經傳；或親身經歷後，相當失望。評比想做到客觀並不容易。以餐廳為例，評分者可能只熟悉幾間、每個人給分的標準不同，更不要說餐廳的差異性很大，怎麼放在一起比？網際網路時代，常經由網路來評比，卻讓有心人容易操作。種種因素，使評比的可靠性，屢令人存疑。

IMDb(Internet Movie Database)是一關於電影及電視的資料庫，內容極豐富，至 2016 年 6 月，收集的影片約有 370 萬部，註冊會員則多達 6 千 7 百萬位。任一會員，皆可對任一部電影，從 1 至 10 予以評分。IMDb 還依評分高低，提供 250 部佳片的名單。370 萬部電影，分散在 1 百多年間，各影片給分的人又不同，各種評比的困難它都有，因此電影的評分，及 250 佳片名單，有參考價值嗎？

對每部電影，IMDb 並非只給出平均得分，還附上評分總人數、各級分所佔百分比，及評分者之性別、年齡、美國人與非美國人等統計數據。分數雖高，但評分的人很少，是

因宣傳不夠導致知道的人不多，或因此片僅屬少數人的偏愛？評分者年紀偏低，是由於較適合青少年嗎？資訊透明，讓人除分數外，可做各種判斷。新上映的電影，電影公司會不會找人助陣給高分，以刺激票房？恐怕不太會。一方面，通常一部電影的評分人數，可多達幾萬甚至幾十萬，找人作假，杯水車薪，立即被稀釋掉。另一方面，灌水可能沒有好處。要知有些科幻片、打鬥片，或有大明星主演，宣傳做得很大， 一開始的評分可高達 9 點多。這不難理解，因熱門電影，會搶先去看，且還上網評分者，總是這類影片的愛好者。至於那些原本沒興趣，因見到評分高，而被吸引去看者，出了戲院，懊惱上當的人，說不定上網給個低分。一個 1 分、2 分，要好幾個 9 分、8 分才拉得上來。因此長期而言，一部電影會得到它該有的分數，這正是機率裡大數法則的精神。諸如大學排名的起伏，多少也算是反映作假並無好處。如果為了提高排名，過度包裝，吸引一些國外學者或學生前來訪問或就讀，待來了之後，發現根本言過其實，回去後反宣傳，豈不讓排名降低？

250 佳片名單上各電影之評分，與該部電影網頁上所列常略有差異。而且，有些評分很高的電影，卻未名列 250 佳片中。原來要被列入 250 佳片，乃有些條件。較關鍵的有兩點，即評分人數須有 25,000 以上，且只有“定期投票者”(regular voter)，其評分才算有效。評分人數太少之缺點前面已說了，但門檻也不宜設太高，否則將排除掉有內涵而不很大眾化的電影。又經常投票的人，總是較專業且客觀些。在個別電影中，所有評分皆會被計入平均，即尊重每一投票者

之好惡。但在評選 250 佳片時，則設定較一致的標準。另外，250 佳片名單中，各片所列出之評分，乃依下述公式修正：

$$W = R \times \frac{v}{v+m} + C \times \frac{m}{v+m} \quad ,$$

其中 W=估計評分，R=該片之有效平均評分，v=該片之有效評分人數，m=進入 250 佳片的最少評分人數(目前為 25,000 人)，C=全部影片之平均評分(目前為 7.0)。可看出 W 是該片之有效平均評分，與全部影片之平均評分的加權。當電影的有效評分人數愈多，則自己的評分所佔比重便愈大; 反之則小。公式 W，其實源自貝氏統計(Bayesian statistics)。要估計某一參數，依過去經驗，我們對該參數可有事先的(prior)看法。經觀測後產生新資訊，便修正以得事後的(posterior)估計。

不論個別影片之評分，或 250 佳片之評比，IMDb 均成功地將統計的概念引入，使得本該不易評比，卻變成一可信的評比。