

資料收集的誤用

隨著統計用途的日益廣泛，統計被誤用的情形日趨嚴重，有些誤用是無意的，但也有刻意誤用的狀況，為了避免被這些不正確的統計誤導，我們應該要對統計的誤用有清楚的認識。一般而言，統計學的誤用大致可分為以下四個方面：資料收集的誤用、資料顯示的誤用、資料分析的誤用和資料解釋的誤用，此處先來討論資料收集的誤用，當然重點在刻意的誤用。

資料收集的誤用可能來自於：

1、不可比較的資料

美國的海軍徵兵廣告中曾宣稱在海軍服役比一般市民還安全，因為當時西班牙與美國交戰期間，美國海軍的死亡率只有 0.9%，而同一期間紐約市的死亡率則高達 1.6%。難道真的在海軍服役死亡率比紐約市民還低？問題在哪裡？因為此二群體的年齡結構、性別和健康情況都不相同，海軍官兵的死亡者是 18~35 歲的健康男性，而紐約市民的死亡者包括新生嬰兒、老人、醫院的病人和窮困的無家可歸者，所以兩者根本不該比較。

你知道全球 2018 年各國的碳排放量的前三名嗎？第一名是中國，排放量為 10,065 百萬噸 CO₂，佔全球排放量的 27.5%；第二名是美國，共排放 5,416 百萬噸 CO₂，佔全球排放量的 14.8%；第三名是印度，共排放 2,654 百萬噸 CO₂ 佔全球排放量的 7.25%。中國的排放量竟達到美國的近二倍，實為造成極端氣候的元凶！這個說法有問題嗎？

根據 USDA 數據顯示，2019 年全球豬肉消費量達到 10,090 萬噸，中國排名第一，消費了 4,486.6 萬噸，占比高達 49% 左右，而美國只排名第三(9.97%)，中國幾乎是美國的 5 倍，也就是說全世界大約有一半的豬肉是被中國人吃掉的。為了供應牲畜的飼料，亞馬遜雨林不斷減少，另外農藥和殺蟲劑的使用也造成了環境致命的衝擊！難怪有報告指稱，中國人吃一口肉，亞馬遜雨林就冒一陣煙！這個說法有問題嗎？

2、分類基礎不同

2020 年 12 月 4 日美國公佈 11 月的失業率為 6.7%，台灣行政院主計處公佈 2019 年 12 月的失業率為 3.67%，顯然就失業率比較，台灣人

可要比美國人好太多了。其實失業率的計算學問很大，各國計算失業率的基礎和調查的嚴謹度都是不相同的。美國的失業者可以申領救濟金，找不到工作的人登記後才可申領，因此統計數據就是根據申領人數得到的。而台灣並沒有失業保險，在當事人和政府都要「面子」的情況下，許多失業者並不願意讓人知道自己失業，而以回老家幫忙加以隱藏。再者，暫時擺地攤、在補習班補習、做計時工、公家機構的短期契約人員……等算不算失業？你說呢？

新冠疫情仍在全球肆虐，可是關於核酸檢測結果是否為陽性，各國的判定標準也不一樣，於是就有甲國檢測為陰性，到了乙國確成陽性的案件。此外對於是否本土或境外移入、是否納入確診統計、是否納入死亡統計，其實各國標準也不是一致的。再以美國來說，在新冠病毒流行之前流感已經大爆發，美國疾控中心 2020 年 2 月 4 日的報告，自 2019 年 9 月起已導致全美 1900 萬人染病，18 萬人住院、1 萬人死亡，可是當時只有 11 人是新冠肺炎確診病例。之後沒多久，好像流感就消聲匿跡，新冠則爆炸式成長，很快時間就達成了「美國優先、美國第一」的目標。為什麼會這樣？中間沒有玄機嗎？

3、偏差的樣本

當樣本無法正確的代表母群體時，當然會造成很大的錯誤。其中最著名的例子是 1936 年的美國總統大選，*Literary Digest* 雜誌曾預測共和黨候選人 Landon 會大勝民主黨的候選人 Roosevelt。但實際選舉結果揭曉後卻完全相反，Roosevelt 除了在 Maine 和 Vermont 兩州外，在其餘州皆大獲全勝，總共獲得了 60% 的選票，由於這個大烏龍，該雜誌不到一年就關門大吉了。其實該雜誌曾經連續四次正確的預測了美國總統大選的結果，用的方法都是根據電話號碼簿和各種雜誌訂戶名單，寄出模擬選票。在 1936 年那次，甚至還寄出了 1000 萬張模擬選票，回收了約 230 萬份，樣本數不可謂不大，那問題出在哪裡呢？原來當時能夠擁有電話或訂閱雜誌的人，多半是社會中較富有的階級，而傳統上，這些人多半支持共和黨，而羅斯福是民主黨，因此樣本雖大但並不能代表具有投票權的母群體。

好了，以後我們在看台灣選舉民調時，請注意是由哪個民調機構做的，如果是黨派自己做的或民調機構有非常明顯的傾向性時，我們不要忘了，通常也一樣會出現類似的問題。

4、不正確的計算

在 2020 年年底時，主計處公佈最新預測，台灣人均 GDP 將於 2021 年首次突破 3 萬美元達 30,038 美元。請問，你對這個好消息有感嗎？為什麼？因為這種所謂人均 GDP 並不是一個好的指標，可是這個指標已經呼嚨了世人好多年了直到現在！就以台灣來說，2011 年是 20,866 美元、1992 年是 10,768 美元。比起十年前、二十年前，請問你感覺到生活一天天好起來了嗎？甚至好了近 3 倍嗎？什麼！二十年來什麼都漲，只有薪水沒漲？怎麼會這樣？

首先，所得不但不是常態分配，還是非常偏的偏態分配，少數有錢人擁有了大多數的財富，計算平均有什麼用？再者，你知道近 15 年來，台灣的 GDP 定義修改了 4 次，每改一次 GDP 就多了幾千億元。最近，又因為美元貶值，所以 GDP 就突破了 3 萬美元！

這類不正確的計算生活上還有很多。例如：股價指數，眼看著大盤大漲，可是你的股票偏偏不漲；明明物價漲了那麼多，可是政府公佈的物價指數卻只上漲一點點……等等，你知道是什麼原因嗎？

5、其他

另外還有資料收集方法上的錯誤，如問題設計不當、調查方法不正確、訪問員或受訪者偏差、樣本太小或不隨機……等等，一樣會使收集的資料缺乏對母群體的代表性。