

# 平均數的陷阱

主要參考資料：《人人都要懂的統計思維》，(日)篠原拓也 著，

丁妍妍譯，機械工業出版社，2022。

平均數(值)非常有用、方便，在統計中經常出現，因為它可以用來概括群體的代表性特徵。例如，我們常用人均 GDP 來代表該國人民的富裕程度，但是平均數也存在以下的一些陷阱，是我們必需要了解的：

## 一、平均數易受極端值影響

如統計的群體中有特別的極端值存在，那麼就會極大的影響平均數。例如，郭台銘是台灣的首富，當然年所得是非常高的，自然會把他小學同班同學的平均年所得拉得非常高，使其失去了代表性。

## 二、計算時依據的群體如有偏差，會扭曲平均數

在醫學上，如果要比較不同治療方法的療效時，要特別注意到治療方法的選擇並不是隨機決定的，有些治療方法是原本就是身體底子比較好或不好的才會選用。因此，當樣本群體的基礎不一定相同時，就算是出現一些期待的結果，可以說本來就是可預測的或必然的。

## 三、群體太小時平均數無法代表平均水平

在進行研究時，受試者的群體太小時，比較平均值是不太可靠的。對於大群體來說，分布是隱定的，偏離平均數的數據有限，平均數自然有其一定的代表性。但太小的群體分布極不穩定，平均數自然波動也大，一般來說，樣本數至少要達到 30 個以上才可。例如籃球隊上場的 5 位球員中，某位大個子球員在場時就拉高了該隊的平均身高超過對手，一調下場，同一隊的平均身高立刻就輸給對手，這時平均身高的比較就毫無意義。

## 四、如果群體的分布不是常態而是偏態的，平均數也會失去代表性

不是所有變數都是呈常態分配變化的，當偏離常態愈多，平均數就愈不具代表性。例如徵收的年度所得稅，總是有少數人因為所得特別高而繳了很多稅，用平均數來代表群體特徵就不適合了。在偏態分布下，「中位數」是位於群體中間的數值，「眾數」是出現最多的數值，

有時反而比平均數更適合代表群體特徵。但也要注意，這些值雖然可以彌補平均數的弱點，因缺乏平均數具有的數學性質，在統計實務中很少使用。

總之，我們不要盲目的一味相信統計結果，在思考問題時，試著對常識和前提提出疑問，秉持批判性的態度去看待，是強化「洞察本質能力」的第一步。