

如何區分「正常」和「異常」

資料來源：《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。

到什麼程度是正常，從哪裡開始是異常？如畫出那條線？很容易取決於主觀判斷。

統計學的理論告訴了我們如何客觀判斷的方法。學好這些方法，有助於提高自己靈活思考的能力。當我們收集了許多原始資料(數據)，需要畫出分布圖、計算平均數和標準差(表示數據的分散程度)，這是掌握總體特徵的第一步。

在一群數據中，難免會發現有相對其他數據明顯較大或較小的數值(或稱為離群值)，這時我們要考慮是否應該將其作為異常值加以剔除，我們並不能依負責人主觀的決定界線來切，因為換了人，離群值的判定就有可能也跟著變了。所以，需要有客觀的方法來判定，統計學界客觀的處理方法如下：

- 平均數和標準差法

先排除問題數據，然後根據剩餘數據計算平均數和標準差，以平均數加減 3 個標準差作為界線，凡是超過的就判定其為離群值，此法適用於大致符合常態分配的數據群，總數宜大於等於 30 個以上。如果數是太少，平均數較不穩定，就很難判斷群值，這是個難點。

- 四分位數法

先把從大到小的數據依序排好，確定位於全部 $1/4$ 處，稱為第一四分位數(較小四分位數)和 $3/4$ 處，稱為第三四分位數(較大四分位數)的數據。再把這二個四分位數之差的 1.5 倍加到較大四分位數，大於這一數值的就判定為離群值；同樣的，用較小四分位數減去差的 1.5 倍，小於此數值的就判定為離群值。

以身高為例，收集一大群成年人的身高數據，假設得到較大四分位數為 180 公分，較小四分位數為 170 公分，兩者相差 10 公分，其 1.5 倍為 15 公分，因此超過 195 公分和低於 155 公分判定為離群值。應用此法時，當數據集中在中間時，兩個四分位數的差會變小，採用這種方法的話，會出現較多離群值。

採以上二種方法也有太過死板的問題。我們也可試著畫一下數據的分布圖，以看圖形方式來判斷，也不失是一個很好的方法，這是利用在群體中的相對關係來判斷，也就是將「距離」這個概重新定義為相對的東西。

我們有時會發現「理所當然的事情，其實並非如此」，正所謂「常識也是可以懷疑的」，我們有時也有必要大膽重新審視常識，或許可以在那裡啟發出新的想法！

討論問題：

- 請上網查查，目前醫界判定是否「高血壓」的界限是如何訂出來的？應該怎麼判定比較妥當？