

全距、百分位數

一組數據的集中趨勢測定數我們可以把它想成就是指這組數據中可以作為代表性的數據，例如平均是多少、哪個最多、最中間的是哪一個……等。而單單以集中趨勢來代表一組數據的特性仍然不夠充分，例如在沙漠地帶日間可達 3~40°C 熱得半死，夜間又可降到 0°C 以下還要蓋大棉被，如果說平均氣溫 20°C，那只是假象完全不能反映實際情況，所以還需要知道這組數據分布的情形。用來顯示數據離開中間數值散布情況的測定數就稱為離中趨勢測定數 (Measure of Dispersion) 或離差、離度，同樣的，我們可以把它想成是這組數據分散(佈)狀況的指標。用專業術語來說，就是指描述一組數據離散程度的量數，也可以稱為「差異量數(Measures of Variability)」，常用的離中趨勢測定數有全距、百分位數、平均差、平方和、變異數、標準差、……等等。以下先就全距和百分位數說明之。：

一、全距(Range)

全距就是一組數據中最大值和最小值之間的差距，常以 R 符號表示。全距就是所有數據的變動範圍，在此範圍內包含了所有的觀察值。全距取決於該組數據中的最大值和最小值，對於中間數據間的差異狀況是無法顯示的，所以二組數據雖然全距相同，但是可能中間數值差異頗大，這是全距無法顯示的，是它的主要缺點。

全距是最大值減去最最小值得到的差距，所以是一個單一數值，有時，我們也可以用「 $X_{\max} \sim X_{\min}$ 」的方式來表達。在日常生活當中，用這種表達方式的例子並不少，例如每天都有的氣象報告(最高氣溫和最低氣溫)、大學聯招放榜時的各校系錄取分數(最高分和最低分)都是很好的例子。

二、百分位數 (Percentiles)

「百分位數」也可以算是一種離中趨勢測定數。百分位數要將所有數據按大小順序排列，排在百分之多少的就是多少百分位數。最常用的有第 5、第 25、第 50、第 75、第 95 百分位數。它的涵意是所有數據中低於它的百分比。例如女性身高的第 5 百分位數(5th percentile)若為 150 公分，就意味著只有 5% 的女性身高矮於 150 公分。

百分位數不但可以用來表示一群數據集中的情況(即第 50 百分位

數)，亦可用來表示該群數據的散佈範圍(即第 0 百分位數到第 100 百分位數)及某個別數據在群體中的相對位置，如由此點來看，它就屬於離中趨勢測定數。

所謂的中位數其實就是第 50 百分位數，亦即將數據依序排列後等分為二個部份，位於中間的數值，所以也可稱為二分位數。

同理，如果分為四個等分，那麼第 25 百分位數就是四等分的第一個分割點，也可稱為第一四分位數(Q1)；第 50 百分位數就是第二四分位數(Q2)，也相當於中位數)；第 75 百分位數就是第三四分位數(Q3)。第三四分位數與第一四分位數的差距又稱為四分位距(Inter Quartile Range, IQR)，這也是另一種離中趨勢測定數。

也有人用五數綜合(five-number summary)來表示離度，所謂五數是指最小數、第一四分位數、中位數、第三四分位數及最大數，從小寫到大，依據這五數畫出的圖形稱為「盒圖」(boxplot)，如果把好幾組數據都畫成盒圖來相互比較也是一個很有效的方法。

分成十等分時，則可以求得第一十分位數至第十十分位數，可以符號 $D_1 \sim D_{10}$ 表示之。

百分位數在人性化設計的運用時是一項常用的指標，例如「最大母群體值」最常採用的就是人體測計數據庫的第 95 百分位數(95th percentile)，「最小母群體值」就是第 5 百分位數(5th percentile)。

百分位數在大學入學的指考中也有運用，所謂的指考五標，即為：

- 頂標：該科成績位於第 88 百分位數之考生級分
- 前標：該科成績位於第 75 百分位數之考生級分
- 均標：該科成績位於第 50 百分位數之考生級分
- 後標：該科成績位於第 25 百分位數之考生級分
- 底標：該科成績位於第 12 百分位數之考生級分

如果是常態分配，那麼第 50 百分位數、中位數和平均數都應該是相等的。