變 異 數

變異數也是一種離中趨勢測定數。

前面提到平方和會受到數據個數的影響，當觀測值數目增加時，平方和會隨個數的增加而增加，因此將平方和除以數據的個數得到的平均平方和，由於此量數有許多用處，我們特別稱之為「變異數」(Variance)，用符號 V 來表示。通常母群體的變異數我們用 V 來表示，而樣本的變異數就用 s^2 來表示。變異數在統計分析中是常用的離差之一，尤其是在統計推論中更為重要。

變異數即為平均之平方和。在作統計推論時，母群體之平均數 μ 是未知的，其變異數自然無法求得，所以只好用樣本的變異數來估計。只是由樣本求出的樣本變異數 s^2 不是真正的母群體變異數 V，其間具有偏差(偏低)，而且樣本數愈少，低估愈嚴重。茲以下述二種方式說明低估的原因：

1. 由於樣本僅為母群體的一部分，以全距來看，除非樣本組中同時抽到母群體中的極大值與極小值，否則樣本組的全距會較母群體的全距偏低。當樣本數愈小時，同時抽到兩個極端值的機會也愈小，偏低的程度也愈多。變異數與全距同為表示離差的測定數，樣本變異數比母群體變異數偏低的道理是一樣的。

2. 在計算樣本變異數時，係以樣本平均數 $\bar{x}$ 代替母群體平均數 μ，除非樣本平均數 $\bar{x}$ 剛好等於 μ，否則 $\sum(xi-\bar{x})^2$ 必然小於$\sum(xi-\mu)^2$。因為 $\bar{x}$ 是由樣本的 x_i 計算出來的，因此 $\sum(xi-\bar{x})$ 必為 0，真正的母群體平均數 μ ，不論是大於或小於 $\bar{x}$ ，所求出的 $\sum(xi-\mu)$ 自然不會為 0。若先將 $(xi-\mu)$ 平方以後再相加，與 0 的差值會更大。因為樣本變異數係以 $\sum(xi-\bar{x})^2/n$ 計算，故會比 $\sum(xi-\mu)^2/N$ 為小，故要以樣本變異數代替母群體變異數就會有偏低的現象。

為了糾正樣本變異數估計母群體變異數時的低估現象，只要在計算樣本變異數時，將分母由 n 改為 n – 1 即可。 此 n – 1 稱為自由度。如此求出來的樣本變異數 s^2 稱為母群體變異數 σ^2 的不偏估計

值。也就是說，此時之 s^2 稱為不偏變異數，此「不偏」係指用它估計母群體變異數時不會偏低。不偏變異數的另一個計算公式如下：

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}$$

由不偏變異數的計算公式中可以看出，當樣本數 n 愈小時，n – 1 對偏差糾正的程度愈多，例如 n = 10 時，平方和原來是除以 10，現在變成除以 9 ，n = 5 時，由除以 5 變成除以 4，差得更多，若 n = 100 時，n – 1 = 99 ，差別就有限了。一般在實務上，當樣本數大於 30 以上時稱為大樣本，分母的減 1 就可以省略。