變異係數

如果有多組不同單位的數據要比較其離差就會比較困擾，例如下列三組資料：

A 組： 100，200，300，400 ，500 $s_A = 100\sqrt{2}$

B 組： 1，2，3，4，5 $s_B = \sqrt{2}$

C 組： 11，12，13，14，15 $s_C = \sqrt{2}$

因為沒有單位，所以 $s_A = 100\, s_B$ ， $s_B = s_C$ 。若假設 A 組的單位為公分，B 組的單位為公尺，C 組的單位為公斤，那麼 $s_A = 100\, s_B$ 看來就不正確了，因為 $100\sqrt{2}$ 公分與 $\sqrt{2}$ 公尺應該相等，而 $s_B = s_C$ 就變成 $\sqrt{2}$ 公尺等於 $\sqrt{2}$ 公斤，好像並不能代表什麼意義。

再假設把 B 組的單位改成公斤，A、C 組不變，則 s_A 自然不再等於 s_B ，但是否等於 s_C 呢？我們先看 B 組是由 1 公斤至 5 公斤，最大變化量為 4 公斤，是最小值 1 公斤的 4 倍，而 C 組由 11 公斤變化至 15 公斤，最大變化也是 4 公斤，但不過是最小值 11 公斤的 4/11 倍，顯然，B 和 C 兩組數據間的變異程度是不同的。

為了解決以上問題，我們可以將標準差除以本身的平均數，也就是算出標準差佔平均數的相對比率(變化量)。由於標準差和平均數的單位是一樣的，相除後單位就互相抵銷，變成一個比值(百分比)，我們就稱為「變異係數」(Coefficient of Variance)。因為變異係數常以 % 表示，也常簡稱為 CV %。
亦即：

$$CV\,\% = (s / \bar{X}) \times 100\,\%$$
$$\text{或} = (\sigma/\mu) \times 100\,\%$$

變異係數應用在下列的情況下特別有其意義：

1. 比較單位不同時各組資料的變異程度

如上例： 設 A 組單位為公分，B 組單位為公尺

$CV_A\ \% \;=\; s_A / \overline{X}_A \;=\; 100\sqrt{2} / 300 \;=\; \sqrt{2} / 3$

$CV_B\ \% \;=\; s_B / \overline{X}_B \;=\; \sqrt{2} / 3$

故 $CV_A\ \% = CV_B\ \%$

2. 比較單位雖相同，但平均數不同時各組資料的變異程度

如上例：設 B 組單位為公斤，C 組亦為公斤

$CV_B\ \% \;=\; s_B / \overline{X}_B \;=\; \sqrt{2} / 3$

$CV_C\ \% \;=\; s_C / \overline{X}_C \;=\; \sqrt{2} / 13$

故 $CV_B\ \% > CV_C\ \%$

3. 可判斷特殊變異的可能性

根據一般經驗，若 CV % 超過 35 % 時，表示標準差相對於平均數的變化很大，即可能含有異常的極端值。此時，若以標準差作為集中趨勢的代表值，其代表性甚差，因為在這種情況下，分配的型態多半不會是單峰對稱型。又若 CV % 小於 5 % 時，顯示變異的幅度過小，有可能抽樣的方法出了問題(樣本經過篩選)或控制過於嚴格所致，此時，應注意分析其造成的原因。