Leonardo Latella

NOTA SULLA DEVIAZIONE STANDARD

Per mezzo degli scarti si possono definire vari tipi di errore da usare come incertezza sulla singola misurazione assumendo $\overline{x}$ come presunto valore dalla grandezza. Uno di questi scarti è la deviazione standard. Vi è differenza tra la deviazione standard di un campione e quella di una popolazione che, prevedendo un numero di misurazioni infinito, non può essere calcolato ma soltanto stimato. L'espressione della deviazione standard è

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}{N}}$$

Questa relazione tende a sottostimare le incertezze nel caso in cui si ha a che fare con pochi valori ($N < 30$) per questo una migliore stima della deviazione standard è data da (*)

$$s = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}{N-1}} \simeq \sigma$$

Per valori di N grandi la differenza tra le due espressioni diventa poco significativa.
Questo tipo di scarto indica che il 68,3% di tutte le misurazioni cadono nell'intervallo compreso tra $\overline{x} - \sigma$ e $\overline{x} + \sigma$.
Il quadrato della deviazione standard è chiamato varianza e ha espressione

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 \simeq \sigma^2$$

Come si ottiene l'espressione (*)?

Consideriamo la seguente relazione

$$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x^*\right)^2 \quad \text{con } x^* = \text{ valore vero}$$

dato che $$\left(x_i - x^*\right)^2 = \left[\left(x_i - \overline{x}\right) + \left(\overline{x} - x^*\right)\right]^2$$

otteniamo la relazione $$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x^*\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\left(x_i - \overline{x}\right) + \left(\overline{x} - x^*\right)\right]^2$$

$$= \frac{1}{N}\left[\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 + \sum_{i=1}^{N}\left(\overline{x} - x^*\right)^2 + 2\left(\overline{x} - x^*\right)\sum_{i=1}^{N}\left(x_i - \overline{x}\right)\right]$$

dato che $$\sum_{i=1}^{N}\left(x_i - \overline{x}\right) = 0$$

otteniamo $$= \frac{1}{N}\left[\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 + N\left(\bar{x} - x*\right)^2\right] = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 + \left(\bar{x} - x*\right)^2$$

ma $$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 = s^2$$

quindi $$\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x*\right)^2 = s^2 + \left(\bar{x} - x*\right)^2$$

da cui $$s^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - x*\right)^2 - \left(\bar{x} - x*\right)^2$$

Considerando i valori medi di entrambe i membri otteniamo

$$E\left(s^2\right) = \sigma^2 - \frac{\sigma^2}{N}$$

Il valore medio di $\left(\bar{x} - x*\right)^2$ è per definizione la varianza della variabile $\bar{x}$ $\left(\sigma_{\bar{x}}^2\right)$; questa a sua volta è pari alla varianza della media al quadrato $\left(\frac{\sigma^2}{N}\right)$. Si ottiene quindi

$$E\left(s^2\right) = \sigma^2\left(1 - \frac{1}{N}\right) = \frac{N-1}{N}\sigma^2$$

Ciò vuol dire che mediamente la varianza di un campione di N misure è $\frac{N-1}{N}$ volte quella della popolazione. La stima corretta di σ è dunque

$$\sigma^2 \simeq \frac{N}{N-1}s^2 = \frac{N}{N-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1}$$

da cui $$\sigma \simeq \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1}}$$

Considerando il caso limite dove $N = 1$, cioè quando effettuiamo una sola misura: la prima definizione da il risultato, non molto ragionevole, $\sigma = 0$; la nuova definizione ci da un risultato non definito del tipo $\frac{0}{0}$, rispecchiando così la nostra totale ignoranza riguardo l'incertezza su una singola misura.