Indici di dispersione

Leonardo Bizzoni

March 8, 2024

Fissiamo un insieme di dati $x_1, x_2, \ldots, x_N$, e ne calcoliamo la media $\overline{x}$. Consideriamo l'**insieme degli scarti** al <u>quadrato</u> $(x_i - \overline{x})^2$ con $i \in [1, N]$.

1 Varianza campionaria

$$S^2 := \frac{1}{N-1} * \sum_{i=1}^{N} (x_i - \overline{x})^2$$

Formula alternativa per gli esercizi:

$$S^2 := \frac{1}{N-1} * (\sum_{i=1}^{N} {x_i}^2 - N\overline{x}^2)$$

La **varianza** campionaria S^2 rappresenta la misura della **dispersione** dei dati rispetto alla media.

Se x_i sono espressi in unità di misura (es. metri) S^2 è espresso nel quadrato dell'unità di misura (es. metri quadri). Per ottenere una statistica omogenea si definisce la deviazione standard.

2 Deviazione standard (campionaria)

$$S := \sqrt{S^2} = \sqrt{\frac{1}{N-1} * \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

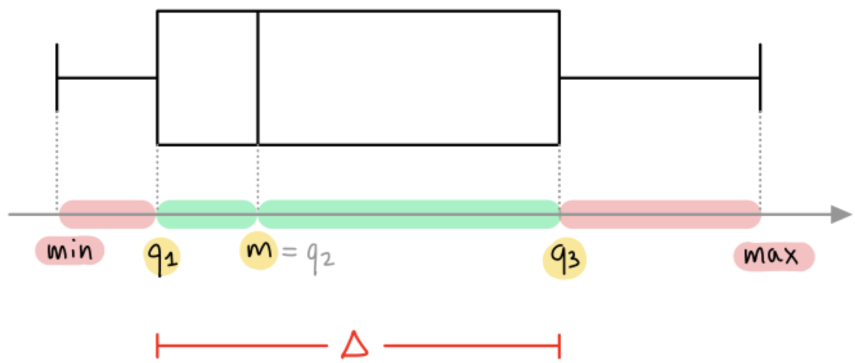
La deviazione standard è la radice quadrata della varianza campionaria, anch'essa misura la dispersione dei dati rispetto alla media ma espressa nella stessa unità di misura, se presente.

3 Scarto interquartile

Un altro indicatore di variabilità, che misura la dispersione dei dati rispetto alla mediana m, è la differenza tra primo e terzo quartile.

$$\Delta := q_3 - q_1$$

Per definizione questo intervallo contiene almeno il 50% dei dati.



4 Osservazione sugli scarti

La somma di tutti gli scarti è sempre nulla:

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

5 Teorema sulla dispersione dei dati

L'intervallo attorno alla media $\overline{x}$ di ampiezza proposizionale alla deviazione S $(\overline{x} - c * S, \overline{x} + c * S)$ con $c \in \mathbb{N} > 1$ contiene una frazione $\alpha \geq 1 - \frac{1}{c^2}$ di dati.

5.1 Esempio

Per $c = 2$ abbiamo che $(\overline{x} - 2 * S, \overline{x} + 2 * S)$ contiene **almeno** il $1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$ dei dati.

Per $c = 3$ abbiamo che $(\overline{x} - 3 * S, \overline{x} + 3 * S)$ contiene **almeno** il $1 - \frac{1}{3^2} = \frac{8}{9} = 88.\overline{8}\%$ dei dati.