Indice di dispersione

Monday, 20 March 2023 15:18

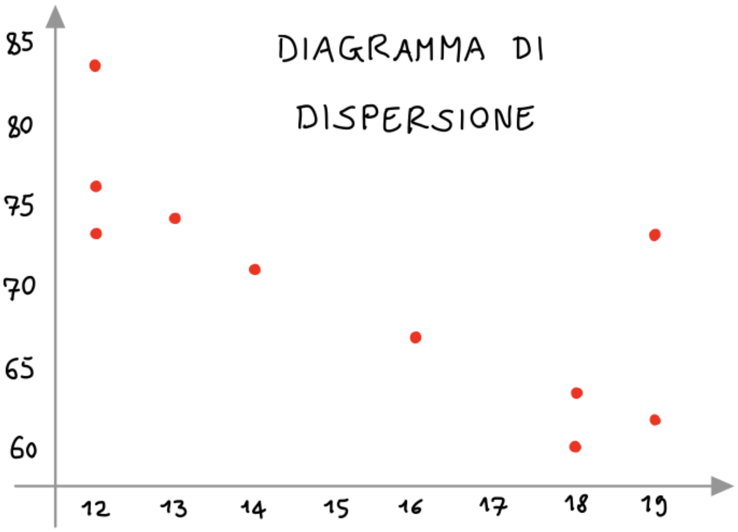
- Ci dice quale tipologia di correlazione i nostri dati hanno
  La formula è:

  $$S^2 := \frac{1}{N-1}\sum(x_i - \bar{x})^2 = \frac{1}{N-1} * \left(\sum(x_i^2) - N * \bar{x}^2\right)$$
  $$= \frac{1}{N-1}\sum x_i - \frac{N}{N-1} \quad * \bar{x}^2$$

  Questa si chiama varianza campionaria
  Che è la media dei nostri scarti
  Abbiamo anche la deviazione standard che non è altro
  $s := \sqrt{s^2}$
  E questo misura la dispersione dei dati rispetto alla media $\bar{x}$
  E quindi, è possibile fare ciò che avevamo fatto con $\Delta = q_3 - q_1$
  Con la deviazione standard con questa formula:
  $(\bar{x} - cs, \bar{x} + cs), c \in R \geq 1$
  Ed a seconda di quanto prendiamo c prendiamo più o meno grande un sottoinsieme
  $c = 2 \rightarrow 75\%, c = 3 \rightarrow 89\%$

  Nota: Varianza si annulla quando tutti i valori sono uguali

- Ora abbiamo dinuovo quell'esempio di prima con i dati in coppia



E' possibile comprendere se una relazione esiste attraverso

E' possibile comprendere se una relazione esiste attraverso
Il coefficiente di correlazione lineare

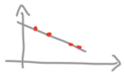
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y} = \frac{\sum x_i y_i - N * \bar{x} * \bar{y}}{(N-1)S_x S_y}$$

E questo ci ritornerà $-1 \leq r \leq +1$

>0

Questo valore:

<0



- R>0 -> corr positiva
- R<0 -> corr negativa
- $|r| \gtrsim 0.7 \rightarrow$ corr. Significativa
- $|r| \lesssim 0.3 \rightarrow$ corr. debole

Es.

(12, 73) (16, 67) (13, 74) (18, 63) (19, 73)

(12, 84) (18, 60) (19, 62) (12, 76) (14, 71)

Noi vogliamo calcolare

$$r = \frac{\sum x_i y_i - N * \bar{x} * \bar{y}}{(N-1)S_x S_y}$$

$$\bar{x} = \frac{1}{N}\sum x_i = \frac{1}{10} * (12 + 16 + 13 + \cdots) = \frac{153}{10} = 15{,}3$$

$$\bar{y} = \cdots = 70{,}3$$

$$S_x^2 = \frac{1}{N-1} * \sum(x_i - \bar{x})^2 = \frac{1}{N-1} * \left(\sum(x_i^2) - N(\bar{x})^2\right)$$

$$= \frac{1}{9} * \left((12^2 + 16^2 + 13^2 + 18^2 \ldots) - 10 * (15{,}3)^2\right) = 9.12$$

$$S_x = \sqrt{S_x^2} = 3.02$$

$$S_y = \cdots = 7{,}36$$

$$\sum x_i y_i = 10603$$

$$r = \frac{10603 - 10 * 15{,}3 * 70{,}3}{(10-1) * 3.02 * 7.36} = -0.76$$

Quindi abbiamo una correlazione negativa