

# Correlazione

Leonardo Bizzoni

March 8, 2024

Consideriamo un insieme di  $N$  dati vettoriali a coppia:  $[(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)]$ .

Vogliamo quantificare la **correlazione** tra le 2 variabili  $x, y$  ossia la tendenza per cui a valori grandi di  $x$  corrispondono valori grandi (*correlazione positiva*) o valori piccoli (*correlazione negativa*) di  $y$ .

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

Dove  $S_x, S_y$  sono le deviazioni standard delle rispettive variabili.

Per  $r > 0$  si ha correlazione **positiva** e per  $r < 0$  si ha correlazione **negativa**. Inoltre:

- $|r| \geq 0,7$  la correlazione è detta **significativa**.
- $|r| \leq 0,3$  la correlazione è detta **debole**.