



# Introduzione ai metodi statistici per le applicazioni industriali parte 3

Antonio Panico

Department of Engineering for Industrial Systems and Technologies  
University of Parma

18 giugno 2025

# La Correlazione tra Variabili

Cos'è la correlazione?

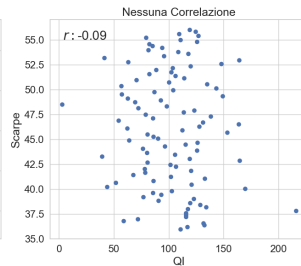
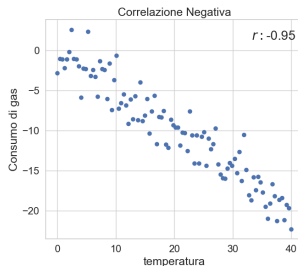
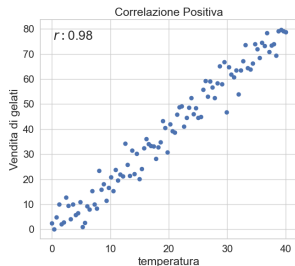
La correlazione misura la relazione lineare tra due variabili. Il coefficiente di correlazione  $r$  assume valori tra -1 e +1:

- $r > 0$ : correlazione positiva
- $r < 0$ : correlazione negativa
- $r \approx 0$ : nessuna correlazione lineare

Formula del coefficiente di correlazione di Pearson

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

# Esempi di correlazione



## Correlazione: un' altra definizione

### Un'altra interpretazione della correlazione

Oltre alla formula, il coefficiente di correlazione di Pearson  $r$  può essere interpretato come un valore che misura quanto due variabili si muovono insieme, rispetto alla loro variabilità individuale

### Formula del coefficiente di correlazione di Pearson:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

- $\text{Cov}(X, Y)$ : covarianza tra  $X$  e  $Y$
- $\sigma_X, \sigma_Y$ : deviazioni standard di  $X$  e  $Y$



# Esempio: calcolo del coefficiente di correlazione $r$

**Dati osservati:**

Osservazione	X	Y
1	1	2
2	2	4
3	3	6
4	4	8



## Esempio: calcolo del coefficiente di correlazione $r$

### Passaggi:

- Media di  $X$ :  $\bar{X} = 2.5$ ,    Media di  $Y$ :  $\bar{Y} = 5$
- Deviazione standard:

$$\sigma_X = \sqrt{\frac{(1-2.5)^2 + \dots + (4-2.5)^2}{4}} = 1.118 \quad \sigma_Y = 2.236$$

- Covarianza:

$$\text{Cov}(X, Y) = \frac{1}{4} \sum (X_i - \bar{X})(Y_i - \bar{Y}) = 2.5$$

- Coefficiente di correlazione:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{2.5}{1.118 \cdot 2.236} = 1.00$$

## Correlazione: confronto con l'esempio precedente

Dati già visti

<b>X</b>	1	2	3	4
<b>Y</b>	2	4	6	8
<b>W</b>	10	20	30	40
<b>Z</b>	20	40	60	80

**Correlazioni:**

$$\text{Corr}(X, Y) = 1.00 \quad \text{Corr}(W, Z) = 1.00$$

### Osservazione

Anche se le covarianze erano diverse, la correlazione standardizzata mostra che la relazione lineare tra le variabili è perfetta in entrambi i casi.



# Correlazione $\neq$ Causalità

**Correlazione** indica che due variabili si muovono insieme.

**Causalità** implica che una variabile influenza direttamente l'altra.

**Attenzione: Correlazione non implica causalità!**

## Esempio ironico

“Aumentano le vendite di cravatte, aumenta il PIL.” → *Non è che le cravatte fanno crescere l'economia! C'è una **terza variabile nascosta**: il miglioramento dell'economia stessa.*

## Messaggio chiave

*Una buona visualizzazione può far intuire una relazione; ma servono studi ben progettati per dimostrare un legame causale.*



# Outlier e Correlazione di Pearson

Cosa sono gli outlier?

Gli outlier sono valori estremi che si distaccano in modo marcato dagli altri dati. Possono essere errori, anomalie oppure variazioni naturali.

# Outlier e Correlazione di Pearson

Cosa sono gli outlier?

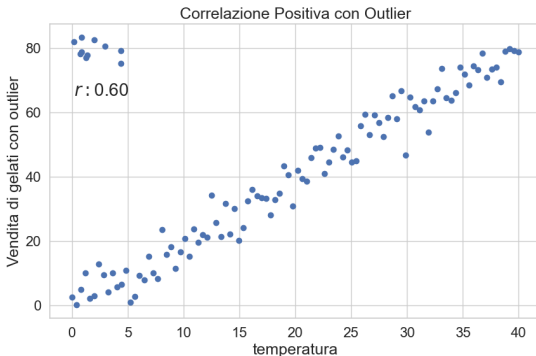
Gli outlier sono valori estremi che si distaccano in modo marcato dagli altri dati. Possono essere errori, anomalie oppure variazioni naturali.



# Outlier e Correlazione di Pearson

## Problema con Pearson

Il coefficiente di Pearson è **sensibile agli outlier**, perché usa direttamente i valori. Un solo punto molto distante può **falsare** il risultato e dare una correlazione ingannevole.



## Calcolo del rango: esempio completo

### Definizione

Il **rango** (o rank) è la **posizione di un valore** quando i dati sono ordinati in modo crescente.

Valori originali (disordinati):

12	8	15	10	9
----	---	----	----	---

Ordinati (per calcolo del rango):

8	9	10	12	15
---	---	----	----	----

1

2

3

4

5

# Calcolo del coefficiente di Spearman

## 1. Cos'è il coefficiente di Spearman?

È una misura di correlazione non parametrica che valuta quanto due variabili siano monotonicamente associate, usando i ranghi invece dei valori originali.

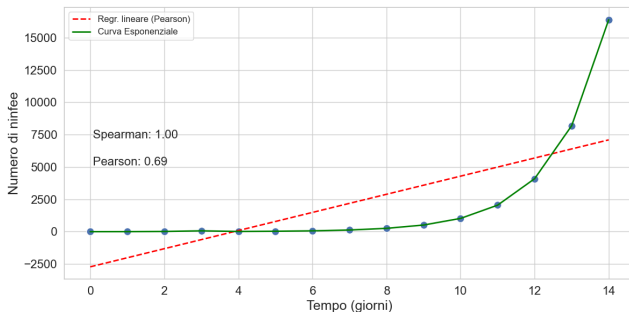
## 2. Passaggi del calcolo

- 1 Assegna un **rango** a ciascun valore di  $X$  e  $Y$
- 2 Calcola la differenza tra i ranghi:  $d_i = R(x_i) - R(y_i)$
- 3 Calcola  $d_i^2$  per ogni coppia
- 4 Applica la formula:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

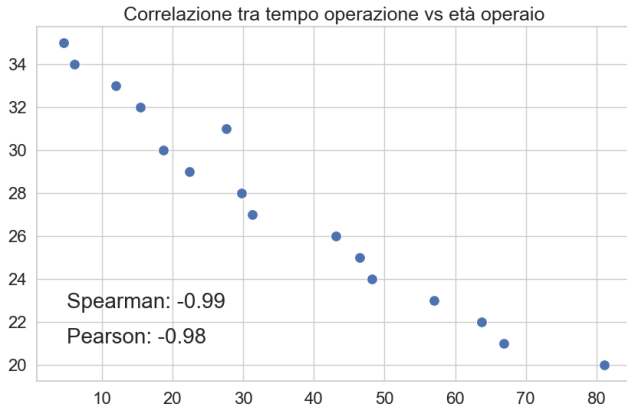
## Nota bene

- I valori devono essere ordinabili
- Se ci sono pareggi, si usano i ranghi medi
- Spearman è **robusto agli outlier** e non richiede relazione lineare



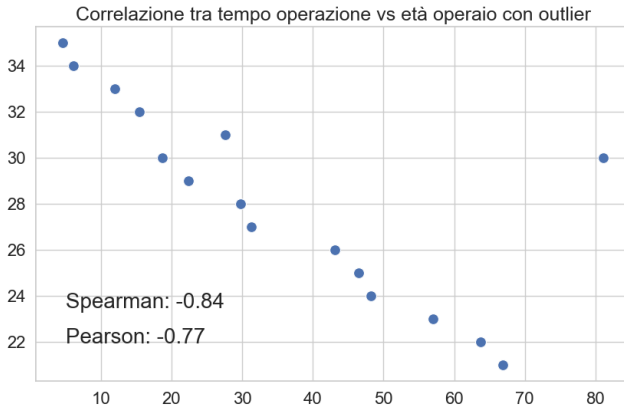
## Nota bene

- I valori devono essere ordinabili
- Se ci sono pareggi, si usano i ranghi medi
- Spearman è **robusto agli outlier** e non richiede relazione lineare



## Nota bene

- I valori devono essere ordinabili
- Se ci sono pareggi, si usano i ranghi medi
- Spearman è **robusto agli outlier** e non richiede relazione lineare





# Cosa sono le Heatmap?

## Definizione

Una **heatmap** (mappa di calore) è una rappresentazione visiva dei dati dove i valori vengono mostrati attraverso variazioni di colore.

# Cosa sono le Heatmap?

## Definizione

Una **heatmap** (mappa di calore) è una rappresentazione visiva dei dati dove i valori vengono mostrati attraverso variazioni di colore.

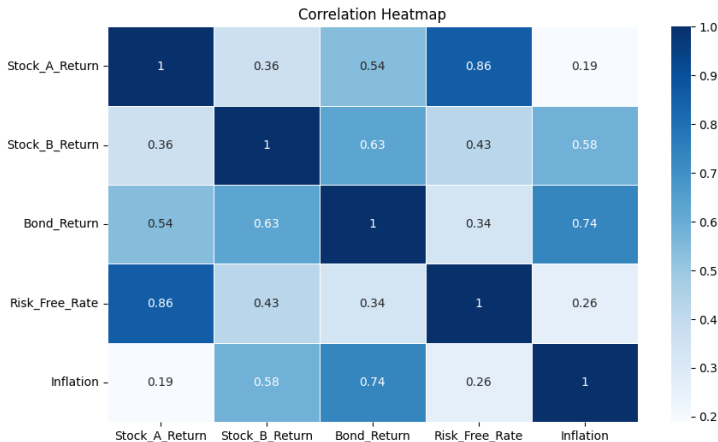
- Ogni cella rappresenta un valore numerico, come una matrice o una tabella.
- I colori permettono di identificare facilmente zone con valori più alti o più bassi.
- Spesso usata per visualizzare:
  - Matrici di correlazione
  - Frequenze o intensità
  - Andamenti su due variabili categoriali o continue

## Esempio di Dataset Finanziario

Periodo	Stock A	Stock B	Bond	Risk-Free	Inflation
1	0.12	0.10	0.03	0.01	0.02
2	0.15	0.12	0.04	0.02	0.03
3	0.08	0.07	0.02	0.01	0.02
4	0.10	0.08	0.05	0.01	0.04
5	0.09	0.14	0.04	0.01	0.03
6	0.16	0.09	0.03	0.02	0.02
7	0.11	0.13	0.06	0.01	0.03
8	0.13	0.16	0.05	0.02	0.04
9	0.07	0.11	0.02	0.01	0.03
10	0.18	0.15	0.07	0.02	0.04

- **Stock A / B:** Rendimenti di due azioni diverse.
- **Bond:** Rendimento obbligazionario.
- **Risk-Free:** Tasso privo di rischio (es. titoli Stato).
- **Inflation:** Tasso d'inflazione nel periodo.

# Heatmap Esempio Finanziario



# Strategia di Ottimizzazione del Portafoglio

## Obiettivo

Massimizzare il rendimento atteso mantenendo il rischio sotto controllo tramite una buona diversificazione.

## Analisi della heatmap di correlazione:

- **Stock A e Risk-Free Rate:** correlazione molto alta ( $r = 0.86$ ) → attenzione alla ridondanza.
- **Stock B e Bond:** correlazione significativa ( $r = 0.63$ ) → possibile sovrapposizione nei movimenti.
- **Bond e Inflazione:** correlazione forte ( $r = 0.74$ ) → rischio inflazione presente.
- **Stock A e B:** correlazione moderata ( $r = 0.36$ ) → buona combinazione per diversificare.

## Nota:

*una buona diversificazione si ottiene evitando asset altamente correlati tra loro.*

È il momento di mettersi alla prova!

Hai seguito tutta la spiegazione? Ora puoi testare la tua comprensione!

**Partecipa al quiz online:**

[Apri il quiz su Google Moduli](#)