

# Esercitazione 2

Corso di Statistica

# Info

Daniele Pasquini

Phd Student in Data Science

- **Email:** [daniele.pasquini@uniroma2.eu](mailto:daniele.pasquini@uniroma2.eu)
- **Ricevimento:** Online su Teams
- **Ricevimento:** Ing. Informatica, Corpo D, Piano 2, Stanza 20

# Mediana

- Valore centrale di un set di dati (ordinati)
- valore che occupa la posizione centrale, se  $n$  è **dispari**
- Media dei due valori centrali, se  $n$  è **pari**
- Nei dati **discreti** , la mediana si trova **guardando la frequenza cumulata.**

# Mediana

- Nei dati in classi, si calcola per interpolazione

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$
$$M_e = L + ((0.5 - F'_c)/f'_m) \cdot h$$

dove:

$L$  = limite inferiore della **classe mediana**

$N$  = totale delle frequenze

$F_c$  = somma delle frequenze **precedenti** la classe mediana

$f_m$  = frequenza della classe mediana

$h$  = ampiezza della classe

# Interpolazione

- L'interpolazione serve per **ricavare un valore "interno"** a un intervallo, quando non lo conosciamo esattamente ma sappiamo come varia la grandezza ai suoi estremi
- Quando i dati sono **raggruppati in classi non conosciamo i valori singoli**, ma solo:
  - L'ampiezza della classe
  - La frequenza di ogni classe
- Stimare (interpolare) il punto preciso dentro la classe che contiene il valore cercato

# Mediana

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$

dove:

$L$  = limite inferiore della **classe mediana**

$N$  = totale delle frequenze

$F_c$  = somma delle frequenze **precedenti** la classe mediana

$f_m$  = frequenza della classe mediana

$h$  = ampiezza della classe

| Classe di età (anni) | Numero di persone $f_i$ | Cumulata |
|----------------------|-------------------------|----------|
| 20 - 25              | 80                      | 80       |
| 25 - 30              | 100                     | 180      |
| 30 - 35              | 90                      | 270      |
| 35 - 40              | 30                      | 300      |
| <b>Totale</b>        | <b>N = 300</b>          |          |

- Totale  $N = 300$
- La **mediana** è il valore che divide il campione a metà, la posizione mediana è:  $N/2 = 150$
- La 150<sup>a</sup> osservazione cade nella **classe 25-30**, perché: fino alla prima classe (20-25) abbiamo 80 osservazioni, fino alla seconda (25-30) arriviamo a 180.
- la 150<sup>a</sup> è dentro la classe 25-30.

# Mediana

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$

dove:

$L$  = limite inferiore della **classe mediana**

$N$  = totale delle frequenze

$F_c$  = somma delle frequenze **precedenti** la classe mediana

$f_m$  = frequenza della classe mediana

$h$  = ampiezza della classe

| Classe di età (anni) | Numero di persone $f_i$ | Cumulata |
|----------------------|-------------------------|----------|
| 20 - 25              | 80                      | 80       |
| 25 - 30              | 100                     | 180      |
| 30 - 35              | 90                      | 270      |
| 35 - 40              | 30                      | 300      |
| <b>Totale</b>        | <b>N = 300</b>          |          |

stimare dove si trova esattamente **dentro** quella classe con interpolazione lineare:

$L$  = limite inferiore della **classe mediana** → 25

$N$  = totale delle frequenze → 300

$F_c$  = somma delle frequenze **precedenti** la classe mediana → 80

$f_m$  = frequenza della classe mediana → 100

$h$  = ampiezza della classe → 30-25 = 5 anni

# Mediana

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$

dove:

$L$  = limite inferiore della **classe mediana**

$N$  = totale delle frequenze

$F_c$  = somma delle frequenze **precedenti** la classe mediana

$f_m$  = frequenza della classe mediana

$h$  = ampiezza della classe

| Classe di età (anni) | Numero di persone $f_i$ | Cumulata |
|----------------------|-------------------------|----------|
| 20 - 25              | 80                      | 80       |
| 25 - 30              | 100                     | 180      |
| 30 - 35              | 90                      | 270      |
| 35 - 40              | 30                      | 300      |
| <b>Totale</b>        | <b>N = 300</b>          |          |

Mediana = 28.5

L'interpolazione serve a "entrare dentro" la classe 25-30 e **stimare** il valore esatto (28,5) in cui si trova la 150<sup>a</sup> persona, **assumendo che i dati siano distribuiti uniformemente** dentro la classe.

# Mediana

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$

$$M_e = L + \left( (0.5 - F'_{prec}) / f'_{med} \right) \cdot h$$

$L$  = limite inferiore della **classe mediana**

$N$  = totale delle frequenze

$F_c$  = somma delle frequenze **precedenti** la classe mediana

$f_m$  = frequenza della classe mediana

$h$  = ampiezza della classe

$F_{prec}$  = somma delle frequenze relative **precedenti** la classe mediana

$f_{med}$  = frequenza relativa della classe mediana

| Classi (n. stanze) | Frequenza $f_i$ | Frequenza relativa $f'_i$ | Cumulata relativa $F'_i$ | Cumulata $F_i$ |
|--------------------|-----------------|---------------------------|--------------------------|----------------|
| 2-3                | 70              | $70/500 = 0,14$           | 0,14                     | 70             |
| 3-4                | 200             | 0,40                      | 0,54                     | 270            |
| 4-5                | 150             | 0,30                      | 0,84                     | 420            |
| 5-6                | 50              | 0,10                      | 0,94                     | 470            |
| 6-7                | 30              | 0,06                      | 1,00                     | 500            |

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Mediana

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$

$$M_e = L + \left( (0.5 - F'_{prec}) / f'_{med} \right) \cdot h$$

**Calcolo della mediana::**

- $L_i = 3$
- $N = 500$
- $F_c = 70$
- $f_m = 200$
- $h = 1$

$$M_e = 3 + ((250 - 70) / 200) * 1 = 3 + (180 / 200) = 3 + 0,9 = \mathbf{3,9}$$

| Classi (n. stanze) | Frequenza $f_i$ | Frequenza relativa $f'_i$ | Cumulata relativa $F'_i$ | Cumulata $F_i$ |
|--------------------|-----------------|---------------------------|--------------------------|----------------|
| 2-3                | 70              | $70/500 = 0,14$           | 0,14                     | 70             |
| 3-4                | 200             | 0,40                      | 0,54                     | 270            |
| 4-5                | 150             | 0,30                      | 0,84                     | 420            |
| 5-6                | 50              | 0,10                      | 0,94                     | 470            |
| 6-7                | 30              | 0,06                      | 1,00                     | 500            |

**APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE**

# Mediana

$$M_e = L + \left( \frac{\frac{N}{2} - F_c}{f_m} \right) \cdot h$$

$$M_e = L + \left( (0,5 - F'_{prec}) / f'_{med} \right) \cdot h$$

**Calcolo della mediana::**

- $L_i = 3$
- $F'_{prec} = 0,14$
- $f'_{med} = 0,40$
- $h = 1$

$$Me = 3 + ((0,5 - 0,14) / 0,40) * 1 = 3 + (0,36 / 0,40) = 3 + 0,9 = \mathbf{3,9}$$

| Classi (n. stanze) | Frequenza $f_i$ | Frequenza relativa $f'_i$ | Cumulata relativa $F'_i$ | Cumulata $F_i$ |
|--------------------|-----------------|---------------------------|--------------------------|----------------|
| 2-3                | 70              | $70/500 = 0,14$           | 0,14                     | 70             |
| 3-4                | 200             | 0,40                      | 0,54                     | 270            |
| 4-5                | 150             | 0,30                      | 0,84                     | 420            |
| 5-6                | 50              | 0,10                      | 0,94                     | 470            |
| 6-7                | 30              | 0,06                      | 1,00                     | 500            |

**APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE**

# Mediana

- **Input, durata in ore di 10 transistor**

[112, 121, 126, 108, 141, 104, 136, 134, 118,  
129]

Ordiniamo i dati:

[104, 108, 112, 118, 121, 126, 129, 134, 136, 141]

# Mediana

| N. stanze ( $x_i$ ) | N. abitazioni<br>(frequenza $f_i$ ) | Frequenza cumulata |
|---------------------|-------------------------------------|--------------------|
| 2                   | 70                                  | 70                 |
| 3                   | 200                                 | 270                |
| 4                   | 150                                 | 420                |
| 5                   | 50                                  | 470                |
| 6                   | 30                                  | 500                |
| <b>Totale</b>       | <b>N = 500</b>                      |                    |

# Mediana

| Superficie (in mq) (Classe) | N. di locali commerciali (Frequenza $f_i$ ) |
|-----------------------------|---------------------------------------------|
| 400 - 1000                  | 351                                         |
| 1000 - 2000                 | 393                                         |
| 2000 - 3000                 | 77                                          |
| 3000 - 5000                 | 39                                          |
| <b>Totale (N)</b>           | <b>860</b>                                  |

# Moda

- valore che si verifica con la maggiore frequenza (cioè che appare più spesso)

# Moda

- **Input, durata in ore di 10 transistor**

[112, 121, 126, 108, 141, 104, 136, 134, 118,  
129]

# Moda

| N. stanze ( $x_i$ ) | N. abitazioni<br>(frequenza $f_i$ ) |
|---------------------|-------------------------------------|
| 2                   | 70                                  |
| 3                   | 200                                 |
| 4                   | 150                                 |
| 5                   | 50                                  |
| 6                   | 30                                  |
| <b>Totale</b>       | <b>N = 500</b>                      |

# Classe modale

| Superficie (in mq) (Classe) | N. di locali commerciali (Frequenza $f_i$ ) |
|-----------------------------|---------------------------------------------|
| 400 - 1000                  | 351                                         |
| 1000 - 2000                 | 393                                         |
| 2000 - 3000                 | 77                                          |
| 3000 - 5000                 | 39                                          |
| <b>Totale (N)</b>           | <b>860</b>                                  |

# Media geometrica

- si usa quando i dati sono **moltiplicativi** (i valori si combinano per prodotto, non per somma)
- utile per **tassi di crescita, rendimenti percentuali, rapporti o serie temporali** dove i valori variano in proporzione.

# Media geometrica

- Se  $x_1, x_2, \dots, x_n$  sono  $n$  valori **positivi**, la **media geometrica** si definisce come:

$$\overline{x_g} = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

$$\overline{x_g} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

# Media geometrica

- Investimento iniziale = 100€
- Primo anno, +10% → moltiplicare per 1.10
- Secondo anno, -10% → moltiplicare per 0.9

$$x_1 = 1,10 \quad \text{e} \quad x_2 = 0,90$$

$$\text{capitale finale} = 100 \times 1,10 \times 0,90 = 99$$

- In due anni abbiamo perso 1€, calo complessivo dell'1%

# Media geometrica

$$\overline{x_g} = \sqrt{1,10 \times 0,90} = \sqrt{0,99} \approx 0,995$$

- Il valore  $\overline{x_g} = 0,995$  rappresenta il **fattore medio di crescita annuale**.
- Significa che, in media, l'investimento si moltiplica ogni anno per **0.995**, cioè **diminuisce dello 0.5%** all'anno.

$$\text{Tasso medio di crescita} = 0,995 - 1 = -0,005 = -0,5\%$$

# Media geometrica

- La media aritmetica dei due tassi (+10% e -10%) sarebbe:

$$\frac{(+10\%) + (-10\%)}{2} = 0\%$$

- Ma il capitale **non resta uguale**: scende da 100 € a 99 €
- la media aritmetica **non descrive correttamente** l'effetto reale nel tempo, mentre la media geometrica sì

# Media geometrica

- Si può calcolare anche con i logaritmi

$$\overline{x_g} = \exp \left( \frac{1}{n} \sum_{i=1}^n \ln x_i \right)$$

# Media geometrica

- Per i dati in classi

$$\overline{x_g} = \left( \prod_{i=1}^k x_i^{f_i} \right)^{\frac{1}{\sum f_i}}$$

Dove:

- $x_i$  = valore discreto / punto medio classe  $i$
- $f_i$  = frequenza di ciascun valore
- $\sum f_i = N$  = numero totale delle osservazioni

# Media geometrica

- Equivalente per evitare esplosione numerica

$$\log(\overline{x_g}) = \frac{\sum f_i \log(x_i)}{\sum f_i}$$

- si eleva 10 o *e* alla potenza del risultato per tornare al valore della media geometrica

# Media geometrica

- Per ogni classe calcola il **punto medio**
- Calcola i **logaritmi** dei punti medi:  $\ln x_i$
- Moltiplica ogni  $\ln x_i$  per la corrispondente frequenza  $f_i$  ottenendo  $f_i \ln x_i$
- Somma i prodotti:  $S = \sum_i f_i \ln x_i$
- Dividi per il totale delle frequenze  $N = \sum_i f_i$ :  $m = S/N$ .
- Prendi l'**esponenziale**:  $\bar{x}_g = e^m$  Se hai usato  $\log_{10}$ , fai  $10^m$

# Media geometrica

| N. stanze     | N. abitazioni  |
|---------------|----------------|
| 2             | 70             |
| 3             | 200            |
| 4             | 150            |
| 5             | 50             |
| 6             | 30             |
| <b>Totale</b> | <b>N = 500</b> |

# Media geometrica

| Superficie (in mq) (Classe) | N. di locali commerciali (Frequenza $f_i$ ) |
|-----------------------------|---------------------------------------------|
| 400 - 1000                  | 351                                         |
| 1000 - 2000                 | 393                                         |
| 2000 - 3000                 | 77                                          |
| 3000 - 5000                 | 39                                          |
| <b>Totale (N)</b>           | <b>860</b>                                  |

$$\log(\overline{x_g}) = \frac{\sum f_i \log(x_i)}{\sum f_i}$$

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Misure di sintesi dei dati

- Le statistiche descrittive si occupano della descrizione e **della sintesi dei dati**
  - **Misure di dispersione**
    - indicano quanto i valori del set di dati sono dispersi o diffusi

# Varianza

- misura della dispersione media dei valori rispetto alla media

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- **media dei quadrati degli scarti** di ogni dato dalla media aritmetica

# Varianza

- per ogni dato, calcoliamo quanto è lontano dalla media, eleviamo questa distanza al quadrato, e poi facciamo la media di tutti questi "quadrati delle distanze".

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

**Nota:** Nella varianza campionaria si utilizza  $n-1$  (dove  $n$  è la dimensione del campione) al denominatore, anziché  $n$ , ed è finalizzata a garantire che la misura ottenuta sia uno **stimatore non distorto** della varianza della popolazione

# Varianza

- Al quadrato perché...

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

**Per eliminare i segni negativi:** La distanza di un punto dalla media può essere positiva (se il punto è sopra la media) o negativa (se è sotto)

# Varianza

- Al quadrato perché...

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- **Per dare più peso agli scarti grandi:** Elevare al quadrato ha l'effetto di "penalizzare" di più i valori che sono molto lontani dalla media (outlier), rendendo la varianza un indicatore sensibile ai dati anomali.

# Esercizio

- **Dati altezze di 5 persone:** 170, 175, 180, 185, 190.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Quale è l'unità di misura finale?

# Standard Deviation

- radice quadrata della varianza campionaria
- utilizzata insieme alla media per quantificare lo spread dei valori

$$s = \sqrt{s^2}$$

- Riporta l'indicatore di dispersione alla **stessa unità di misura dei dati originali**
- Rappresenta la "distanza tipica" o "media" di un'osservazione dalla media del gruppo.

# Esercizio

- **Dati altezze di 5 persone:** 170, 175, 180, 185, 190.

**Quale è l'interpretazione finale?**

# Varianza

- Quando i dati sono raggruppati in classi (es. fasce d'età, classi di reddito), non conosciamo i valori originali
- Usiamo il **valore centrale (mi)** di ogni classe come rappresentante di tutti i dati in quella classe.

$$s^2 = \frac{\sum_{i=1}^n f_i (m_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n f_i (m_i)^2}{n} - \bar{x}^2$$

FORMA DI BASE

FORMA SVILUPPATA

# Varianza e Deviazione standard

| Ore di studio<br>(Classe) | Numero di Studenti<br>(f <sub>i</sub> ) |
|---------------------------|-----------------------------------------|
| 0 - 4                     | 10                                      |
| 4 - 8                     | 20                                      |
| 8 - 12                    | 15                                      |
| 12 - 16                   | 5                                       |
| <b>Totale</b>             | <b>n = 50</b>                           |

$$s^2 = \frac{\sum_{i=1}^n f_i (m_i - \bar{x})^2}{n}$$

$f_i$  frequenza / n.dati di ogni classe

$m_i$  valore centrale di ogni classe

$\bar{x}$  media ponderata:  $1/n * \sum_{i=1}^n f_i(m_i)$

n numero tot osservazioni:  $\sum_{i=1}^n f_i$

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Varianza e Deviazione standard

| Classe        | fi        | mi | fi·mi      |
|---------------|-----------|----|------------|
| 0 - 4         | 10        | 2  | 20         |
| 4 - 8         | 20        | 6  | 120        |
| 8 - 12        | 15        | 10 | 150        |
| 12 - 16       | 5         | 14 | 70         |
| <b>Totale</b> | <b>50</b> |    | <b>360</b> |

$$s^2 = \frac{\sum_{i=1}^n f_i (m_i - \bar{x})^2}{n}$$

$f_i$  frequenza / n.dati di ogni classe  
 $m_i$  valore centrale di ogni classe  
 $\bar{x}$  media ponderata:  $1/n * \sum_{i=1}^n f_i(m_i)$   
n numero tot osservazioni:  $\sum_{i=1}^n f_i$

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Varianza e Deviazione standard

| <b>f<sub>i</sub></b> | <b>m<sub>i</sub></b> | <b>m<sub>i</sub> - x̄</b> | <b>(m<sub>i</sub> - x̄)<sup>2</sup></b> | <b>f<sub>i</sub> · (m<sub>i</sub> - x̄)<sup>2</sup></b> |
|----------------------|----------------------|---------------------------|-----------------------------------------|---------------------------------------------------------|
| 10                   | 2                    | 2 - 7.2 = -5.2            | 27.04                                   | 10 · 27.04 = 270.4                                      |
| 20                   | 6                    | 6 - 7.2 = -1.2            | 1.44                                    | 20 · 1.44 = 28.8                                        |
| 15                   | 10                   | 10 - 7.2 = 2.8            | 7.84                                    | 15 · 7.84 = 117.6                                       |
| 5                    | 14                   | 14 - 7.2 = 6.8            | 46.24                                   | 5 · 46.24 = 231.2                                       |
|                      |                      |                           | <b>Totale</b>                           | <b>648</b>                                              |
|                      |                      |                           | <b>s<sup>2</sup></b>                    | <b>12,96</b>                                            |

$$s^2 = \frac{\sum_{i=1}^n f_i (m_i - \bar{x})^2}{n}$$

$f_i$  frequenza / n.dati di ogni classe  
 $m_i$  valore centrale di ogni classe  
 $\bar{x}$  media ponderata:  $1/n * \sum_{i=1}^n f_i(m_i)$   
 n numero tot osservazioni:  $\sum_{i=1}^n f_i$

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Varianza e Deviazione standard

$$\bar{x}^2 = 7,2^2 = 51,84$$

$$s^2 = \frac{\sum f_i m_i^2}{n} - \bar{x}^2$$

$$m_0 = \frac{4}{2} = 2 \rightarrow m_i^2 \rightarrow 4$$

$$m_1 = 6 \rightarrow 36$$

$$m_2 = 10 \rightarrow 100$$

$$m_3 = 14 \rightarrow 196$$

$$f_i m_i^2 \rightarrow \begin{array}{l} 4 \cdot 10 = 40 \\ 36 \cdot 20 = 720 \\ 100 \cdot 15 = 1500 \\ 196 \cdot 5 = 980 \end{array} \Rightarrow \sum = 3240$$

$$s^2 = \frac{3240}{50} - 51,84 = 12,96$$

Come evidenziato dal risultato...

- La forma base (slide precedente) e la forma semplificata sono **equivalenti**

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Varianza e Deviazione standard

| Class<br>e | f <sub>i</sub> |
|------------|----------------|
| 2          | 70             |
| 3          | 200            |
| 4          | 150            |
| 5          | 50             |
| 6          | 30             |

$$\Sigma f_i x_i = 2 \cdot 70 + 3 \cdot 200 + 4 \cdot 150 + 5 \cdot 50 + 6 \cdot 30 = 1770$$

$$\Rightarrow \bar{x} = 1770/500 = 3.54$$

$$\Sigma f_i x_i^2 = 4 \cdot 70 + 9 \cdot 200 + 16 \cdot 150 + 25 \cdot 50 + 36 \cdot 30 = 6810$$

**Forma sviluppata:**  $s^2 = (6810/500) - (3.54)^2 = 13.62 - 12.5316 = 1.0884 \sim 1.088$

**Forma di base:**  $s^2 = [ \Sigma f_i (x_i - 3.54)^2 ] / 500 = 1.0884 \sim 1.088$  (procedimento slide successiva)

stesso valore, differenze solo se si arrotonda troppo presto

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Varianza e Deviazione standard

| Class<br>e | f <sub>i</sub> |
|------------|----------------|
| 2          | 70             |
| 3          | 200            |
| 4          | 150            |
| 5          | 50             |
| 6          | 30             |

$$\begin{aligned} x_i - \bar{x} &\rightarrow \begin{cases} (2 - 3,54)^2 = 2,3716 \\ (3 - 3,54)^2 = 0,2916 \\ (4 - 3,54)^2 = 0,2116 \\ (5 - 3,54)^2 = 2,1316 \\ (6 - 3,54)^2 = 6,0516 \end{cases} \\ f_i(x_i - \bar{x}) &\rightarrow \begin{cases} 2,3716 \cdot 70 = 166,012 \\ 0,2916 \cdot 200 = 58,32 \\ 0,2116 \cdot 150 = 31,74 \\ 2,1316 \cdot 50 = 106,58 \\ 6,0516 \cdot 30 = 181,548 \end{cases} \rightarrow \Sigma = 544,2 \end{aligned}$$
$$s^2 = \frac{544,2}{500} = 1,0884$$

APPROSSIMATE ALLA TERZA CIFRA DECIMALE SOLO IL RISULTATO FINALE

# Serie storiche

- Una **serie storica** è una sequenza di dati o osservazioni raccolte a intervalli di tempo regolari (ad esempio, giorni, mesi, trimestri, anni) .
- L'obiettivo della sua analisi è comprendere l'evoluzione di un fenomeno nel tempo, individuarne le regolarità e, spesso, fare previsioni per il futuro.

# Serie storiche

- Possiamo scomporre una serie in quattro componenti fondamentali:
  - **Trend:** Rappresenta la tendenza di fondo, il movimento di lungo periodo della serie. Può essere di crescita, decrescita o stabilità.
  - Esempio: il Prodotto Interno Lordo (PIL) di un'economia in via di sviluppo mostrerà un trend crescente su un arco di decenni, nonostante le fluttuazioni di breve periodo

# Serie storiche

- Possiamo scomporre una serie in quattro componenti fondamentali:
  - **Ciclo (Ct)**: Descrive le fluttuazioni ondulatorie di medio periodo attorno al trend.
  - Queste oscillazioni, che durano tipicamente più di un anno, sono spesso legate ai cicli economici di espansione e recessione

# Serie storiche

- Possiamo scomporre una serie in quattro componenti fondamentali:
  - **Stagionalità (St)**: Si riferisce a movimenti e fluttuazioni che si ripetono in modo regolare e prevedibile all'interno di un anno (o di una settimana, o di un giorno).
  - È causata da **fattori climatici** (le vendite di gelati aumentano d'estate) o sociali (le vendite al dettaglio raggiungono un picco a dicembre per le festività natalizie).

# Serie storiche

- Possiamo scomporre una serie in quattro componenti fondamentali:
  - **Componente Irregolare o Accidentale (At)**: Cattura tutte le fluttuazioni residue, casuali e imprevedibili, che non possono essere spiegate dalle altre componenti.
  - Sono dovute a eventi **sporadici** come disastri naturali, scioperi o crisi politiche improvvise

# Esercizio

Un negozio specializzato in articoli per lo sci analizza i suoi dati di vendita mensili dal 2010 al 2025. Osserva che:

- a) Ogni anno, le vendite esplodono da novembre a febbraio, per poi crollare quasi a zero in estate.
- b) Le vendite medie annuali sono aumentate costantemente, grazie alla crescente popolarità della località montana in cui si trova.
- c) Nel febbraio 2021, una nevicata eccezionale e inaspettata ha causato un picco di vendite anomalo e non previsto.
- d) Tra il 2015 e il 2018, c'è stata una fase di crescita accelerata, seguita da un rallentamento fino al 2020, in linea con l'andamento economico generale della regione.

**Domanda:** Associa ciascuna osservazione (a, b, c, d) a una delle quattro componenti della serie storica (Trend, Stagionalità, Ciclo, Irregolarità).

# Esercizio

Una società energetica analizza il consumo orario di energia elettrica in una grande città. Nota che:

- a) Durante i giorni feriali, si verificano due picchi di consumo: uno tra le 8 e le 10 del mattino (quando uffici e fabbriche iniziano l'attività) e uno tra le 18 e le 21 (quando le persone tornano a casa).
  - b) Un pomeriggio di luglio, un violento temporale ha causato un blackout in un quartiere, portando a un crollo improvviso e temporaneo del consumo registrato.
  - c) Negli ultimi 20 anni, il consumo medio giornaliero è in lenta ma costante diminuzione, grazie all'adozione di elettrodomestici a più alta efficienza energetica.
- **Domanda:** Identifica le componenti della serie storica (Trend, Stagionalità, Irregolarità) presenti in queste osservazioni.

# Modelli di composizione

- **Modello Additivo:**

$$Y_t = T_t C + S_t + A_t$$

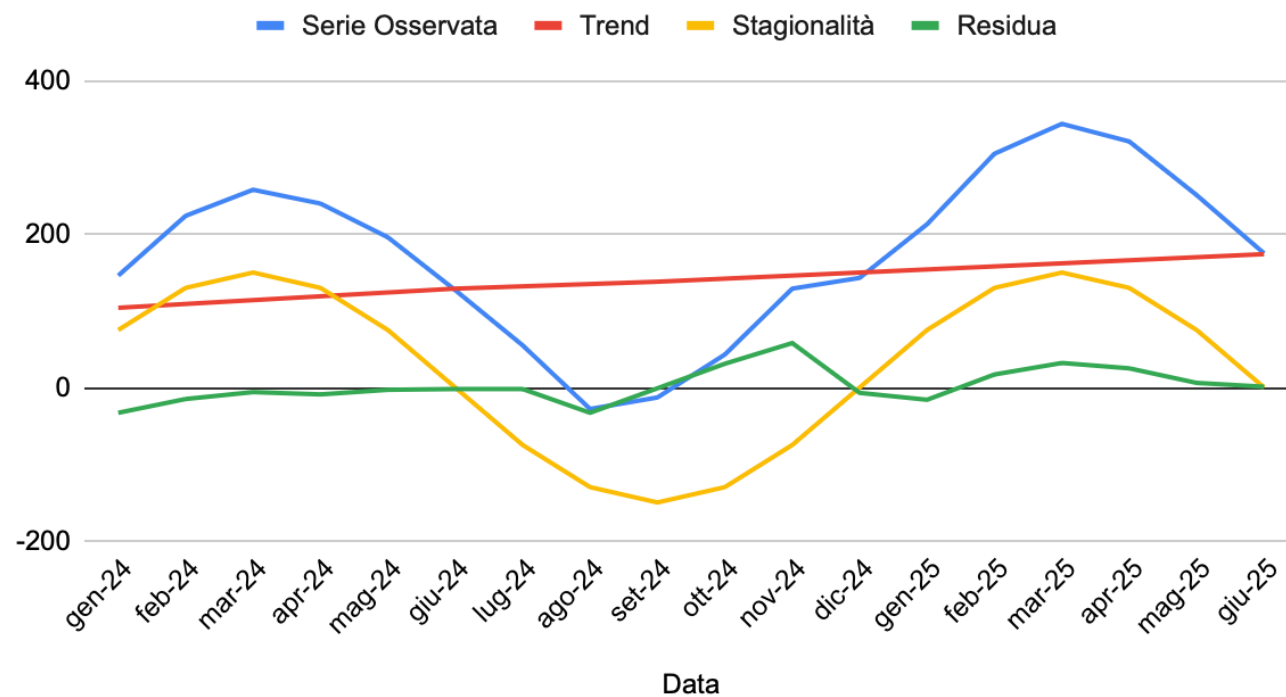
- Si usa quando l'ampiezza delle fluttuazioni stagionali e cicliche è costante nel tempo, indipendentemente dal livello del trend.
- Short: l'effetto stagionale è una quantità **fissa** che si aggiunge o si sottrae al trend

# Modelli di composizione

- **Modello Additivo:**

$$Y_t = T_t C + S_t + A_t$$

Serie Osservata, Trend, Stagionalità e Residua



# Esempio

Chiamate giornaliere ricevute da un call center

- **Trend-Ciclo (CTt)**: A causa di un lento ma costante aumento della popolazione servita, il numero medio di chiamate (la linea di base) aumenta di 10 unità ogni anno.
- Nel 2023 la media è di **500** chiamate/giorno, nel 2024 è di **510**.

# Esempio

Chiamate giornaliere ricevute da un call center

- **Stagionalità (St)**: Esiste un pattern settimanale chiaro:
  - Il lunedì c'è un picco di lavoro, con circa **+80** chiamate rispetto alla media.
  - Il venerdì il numero di chiamate è inferiore, circa **-50** rispetto alla media.

# Esempio

Chiamate giornaliere ricevute da un call center

- **Irregolarità (At)**: Un giorno si verifica un guasto tecnico sulla linea che causa un calo imprevisto di **-25** chiamate.

# Esempio

Chiamate giornaliere ricevute da un call center

- **Calcolo con il Modello Additivo:**
  - **Un lunedì del 2023 senza eventi irregolari:**
    - $Y_t = 500 \text{ (Trend)} + 80 \text{ (Stagionalità)} + 0 \text{ (Irregolarità)} = 580 \text{ chiamate}$
  - **Un lunedì del 2024 con il guasto tecnico:**
    - $Y_t = 510 \text{ (Trend)} + 80 \text{ (Stagionalità)} - 25 \text{ (Irregolarità)} = 565 \text{ chiamate}$

# Esempio

Chiamate giornaliere ricevute da un call center

- **Interpretazione:** l'effetto del "lunedì" è sempre un **incremento fisso di 80 chiamate**, sia che il trend sia 500 o 510.
- L'effetto stagionale non dipende dal livello generale del fenomeno.

# Esercizio

- Un hotel di città ha un numero di prenotazioni medio (trend-ciclo) di 150 camere a notte. A causa di un importante congresso che si tiene ogni anno nella prima settimana di ottobre, le prenotazioni in quel periodo aumentano mediamente di 40 unità. Un martedì di quella settimana, a causa di uno sciopero dei trasporti, si registrano 15 prenotazioni in meno del previsto.
- **Domanda:** Calcola il numero di prenotazioni osservate per quel martedì usando un modello additivo.

# Modelli di composizione

- **Modello Moltiplicativo:**

$$Y_t = T_t C_t \cdot S_t \cdot A_t$$

- È più comune per le serie economiche, dove l'ampiezza della stagionalità tende ad aumentare proporzionalmente al crescere del trend
- **Esempio:** il picco di vendite natalizie è più marcato in valore assoluto negli anni di boom economico
- Se il trend cresce, anche l'impatto (in valore assoluto) della stagionalità cresce

# Modelli di composizione

- **Modello Moltiplicativo:**

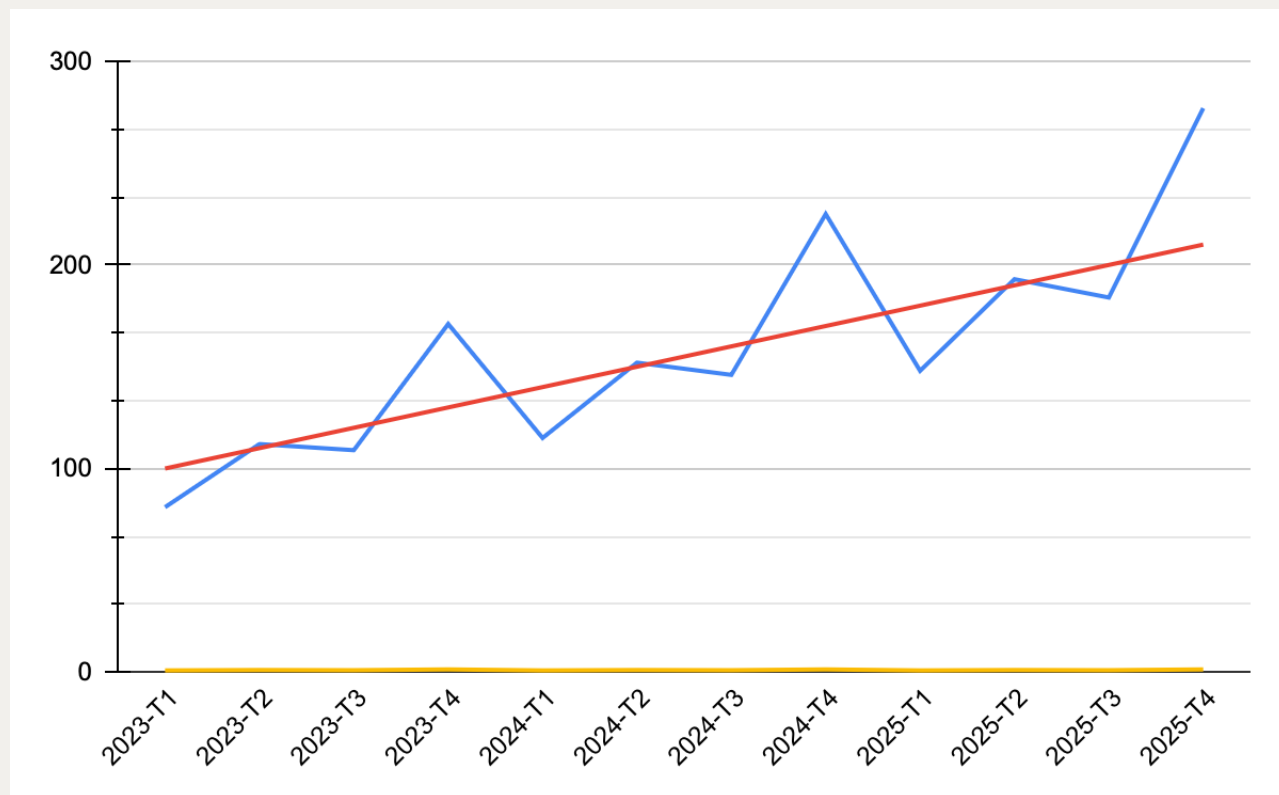
$$Y_t = T_t C \cdot S_t \cdot A_t$$

- le componenti  $S_t$  e  $A_t$  sono espresse come coefficienti o indici (es. 1.20 per un aumento del 20%, 0.90 per un calo del 10%)

# Modelli di composizione

- Modello  
Moltiplicativo:

$$Y_t = T_t C \cdot S_t \cdot A_t$$



# Esempio

Fatturato mensile di una gelateria

- **Trend-Ciclo (CTt)**: L'attività sta crescendo. Il fatturato medio mensile nel 2022 era di **€10.000**. Grazie al successo, nel 2025 il fatturato medio mensile sale a **€25.000**.

# Esempio

Fatturato mensile di una gelateria

- **Stagionalità (St)**: Esiste un forte effetto stagionale:
  - Ad agosto, il fatturato è tipicamente il doppio della media mensile, quindi il fattore stagionale è **1.80** (un aumento dell'80%).
  - A febbraio, il fatturato è solo il 40% della media, quindi il fattore è **0.40**.

# Esempio

Fatturato mensile di una gelateria

- **Stagionalità (St)**: Esiste un forte effetto stagionale:
  - Ad agosto, il fatturato è tipicamente il doppio della media mensile, quindi il fattore stagionale è **1.80** (un aumento dell'80%).
  - Il valore di 1.80, che corrisponde a un aumento dell'80%, è un valore ipotetico scelto per l'esempio

# Esempio

Fatturato mensile di una gelateria

- **Irregolarità (At)**: Un agosto particolarmente soleggiato porta un ulteriore 5% di vendite in più del previsto. Il fattore irregolare è **1.05**

# Esempio

**Come si passa dall'80% al fattore 1.80**

- Nel modello moltiplicativo, ogni componente (stagionalità, irregolarità) agisce "moltiplicando" il valore del trend-ciclo. Il valore del trend-ciclo stesso è considerato la nostra **linea di base, pari al 100% (o 1.0)**.

# Esempio

Come si passa dall'80% al fattore 1.80

- **La Base (Il Trend)**: Il valore del trend rappresenta il 100% del fatturato atteso in assenza di effetti stagionali. In termini di fattore moltiplicativo, questo corrisponde a **1.0**

# Esempio

Come si passa dall'80% al fattore 1.80

- **L'Aumento Stagionale:** Si ipotizza che, storicamente, la gelateria ad agosto venda l'**80% in più** rispetto alla sua media mensile

# Esempio

Come si passa dall'80% al fattore 1.80

- **Calcolo del Fattore Stagionale:** Per ottenere il fattore totale, sommiamo l'aumento percentuale alla base del 100%.
- *Fattore Stagionale = 100% (Base) + 80% (Aumento Stagionale)*  
*= 180%*

# Esempio

Come si passa dall'80% al fattore 1.80

- **Conversione in Numero Decimale:** Per usare questo valore in una moltiplicazione, convertiamo la percentuale in un numero decimale dividendola per 100
- *Fattore Moltiplicativo* =  $100/180 = 1.80$

# Esempio

Come si passa dall'80% al fattore 1.80

- Quando moltiplichiamo il fatturato del trend per **1.80**, stiamo di fatto calcolando il suo valore aumentato dell'**80%**.

# Esempio

Fatturato mensile di una gelateria

- **Calcolo con il Modello Moltiplicativo:**

- **Fatturato di agosto 2022 (senza eventi irregolari):**

$$Y_t = 10.000(Trend) \cdot 1.80(Stagionalita) \cdot 1.00(Irregolarita) = \text{€}18.000$$

- L'effetto stagionale "vale" €8.000 in più rispetto al trend.

# Esempio

Fatturato mensile di una gelateria

- **Calcolo con il Modello Moltiplicativo:**

- Fatturato di agosto 2025 (con l'ondata di caldo):

$$Y_t = 25.000(Trend) \cdot 1.80(Stagionalità) \cdot 1.05(Irregolarità) = €47.250$$

- Qui, l'effetto stagionale "vale" ben €20.000 in più rispetto al trend (€45.000 - €25.000).

# Esempio

Fatturato mensile di una gelateria

- **Interpretazione:** L'aumento del fatturato ad agosto non è una cifra fissa. È un **aumento proporzionale (80%)** che ha un impatto molto più grande in valore assoluto quando il livello generale delle vendite (il trend) è più alto.
- Questo è molto più realistico per fenomeni come il fatturato.

# Esercizio

- Un sito di e-commerce ha un fatturato medio giornaliero (trend-ciclo) di €20.000. Durante il periodo del Black Friday, le vendite sono tipicamente il 250% in più rispetto a un giorno normale. Quest'anno, grazie a una campagna pubblicitaria particolarmente riuscita, le vendite del Black Friday hanno superato le aspettative di un ulteriore 10%.
- **Domanda:** Calcola il fatturato osservato durante il Black Friday di quest'anno usando un modello moltiplicativo.