

# Elementi di Matematica e di Statistica

## Indici di variabilità

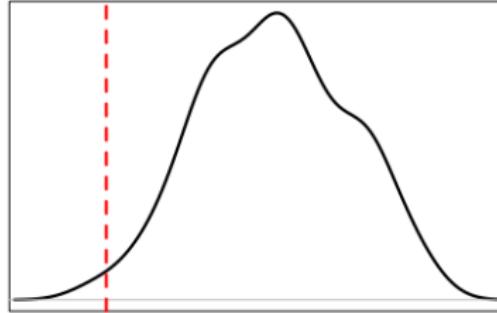
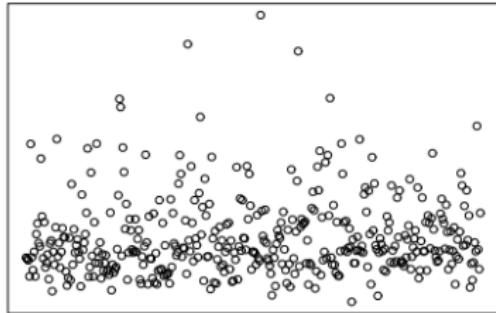
Docente: Riccardo Ievoli  
[riccardo.ievoli@unife.it](mailto:riccardo.ievoli@unife.it)

Corso di Laurea in Biotecnologie  
a.a. 2025-2026

14/10/2025

# Introduzione agli indici di Variabilità

- In realtà la statistica si occupa principalmente di variabilità
- La Poesia di Trilussa:  $\mu_1 = \mu_2 = 1$ ;  $Var_1 = 0 \neq Var_2 = 2$
- Indici che valutano la dispersione attorno alla media
- Indici robusti alla presenza di valori anomali (*outliers*)



# Introduzione agli indici di Variabilità

Indici di *dispersione* attorno alla media

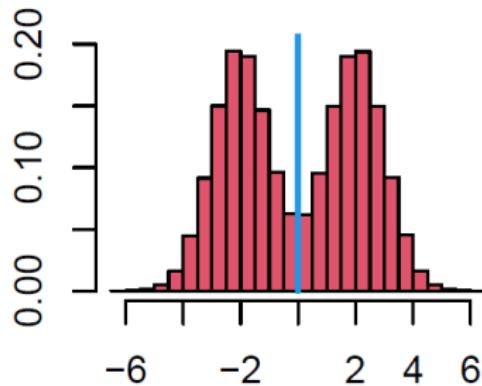
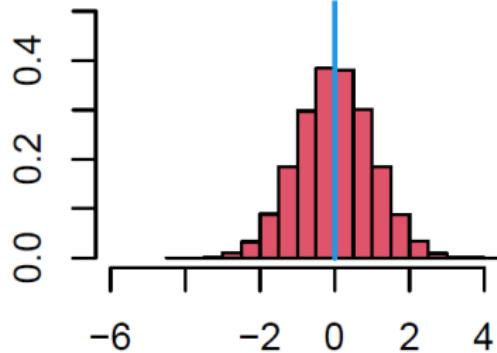
- Devianza e Varianza ( $n\sigma^2; \sigma^2$ )
- Scarto quadratico medio/Deviazione standard ( $\sigma$ )
- Coefficiente di Variazione (CV)
- Problema: La media non è **robusta** (rispetto a cosa?)

Indici di variabilità *robusti* alla presenza di valori anomali

- Range R (Intervallo di variabilità), Differenza Interquartile (IQR)
- Scostamento mediano e deviazione assoluta dalla media

# Misure di dispersione

- L'utilizzo esclusivo di indici di posizione/centralità non è sufficiente.
- Bisogna comprendere *come* le osservazioni si distribuiscono intorno ad un valore centrale.



# Misure di dispersione

**Scopo:** sintetizzare la tendenza di una variabile ad assumere valori distinti nelle unità statistiche

**Punti fermi:**

- se tutte le unità hanno lo stesso valore osservato per la variabile, la misura di variabilità deve essere **nulla**
- se la variabilità dei valori assunti aumenta, la misura di variabilità deve crescere.

**Conseguenza:** le misure di variabilità (univariate) non possono essere negative.

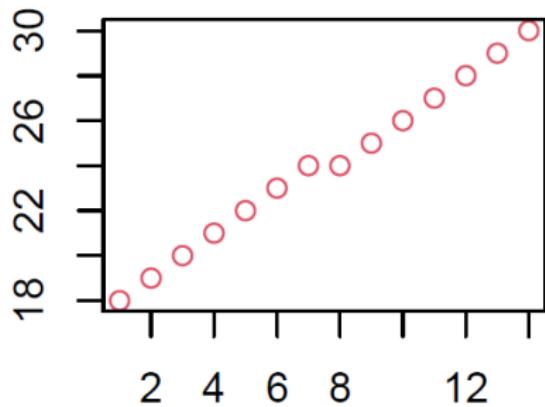
**N.B.:** Generalmente le misure di variabilità riguardano variabili quantitative (ossia numeriche)

# Misure di dispersione

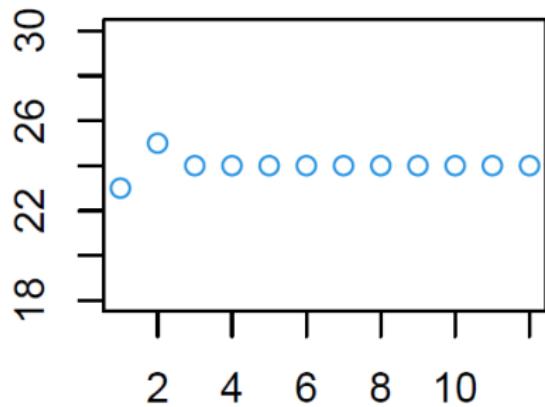
## Esempio 1

- Il voto di due matricole nei primi 14 esami
- La media è pari a 24. E la **variabilità**?
- Quale delle due matricole è stata più variabile nei risultati?

**MATR. A**



**MATR. B**

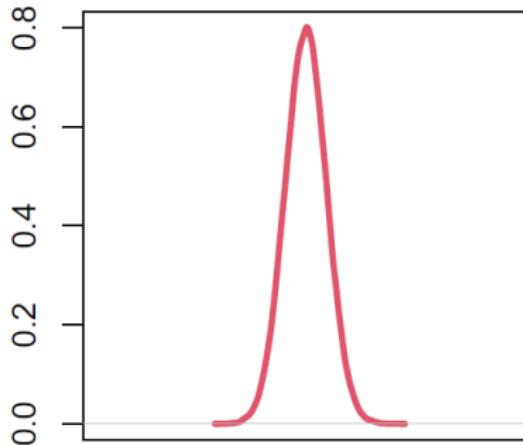


# Misure di dispersione

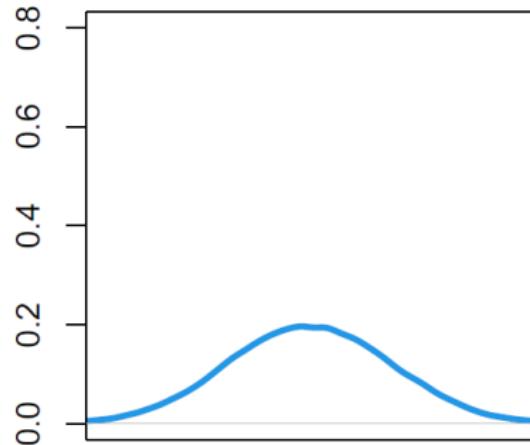
## Esempio 2

- Quale dei due fenomeni è più variabile?

A



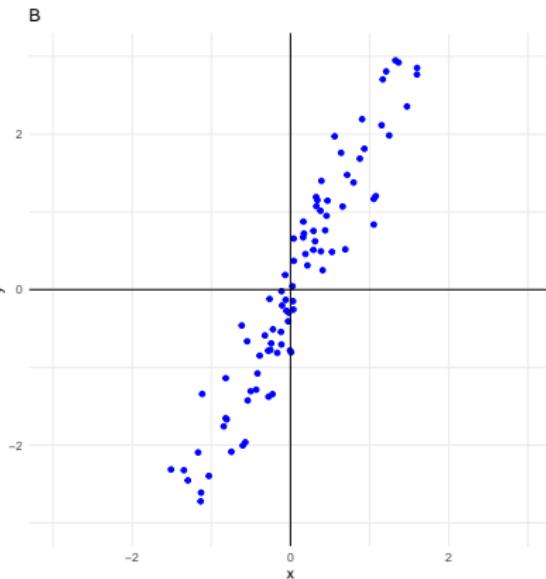
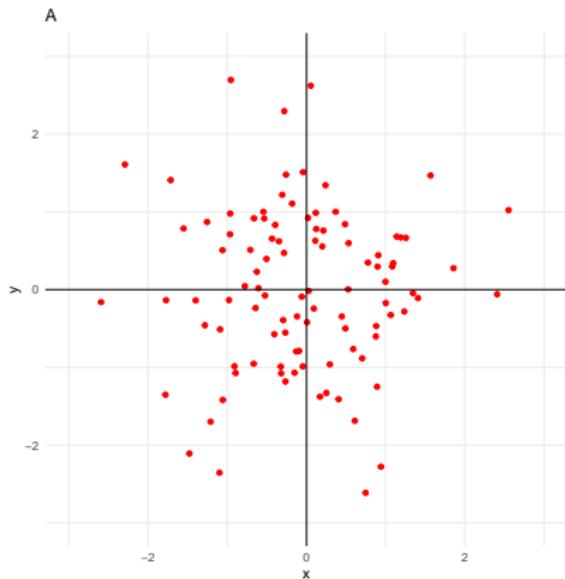
B



# Misure di dispersione

## Esempio 3

- In quale dei due grafici si osserva maggior variabilità?



# Misure di dispersione

## La devianza

- La variabilità viene valutata sulla dispersione delle unità rispetto alla media aritmetica
- La somma degli scarti dalla media al quadrato si chiama **devianza**:

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

# Misure di dispersione

## La varianza

- La devianza non è un indicatore relativo: è legata alla numerosità delle osservazioni ( $n$ )
- La **varianza** risolve questo problema dividendo la devianza per  $n$ : quindi è tecnicamente una media (degli scarti dalla media al quadrato):

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

- A volte la varianza descrittiva è indicata anche con  $\sigma^2$

# Misure di dispersione

## La varianza (2)

- Esiste una formula alternativa per il calcolo della varianza:

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

- In breve, la varianza si può esprimere come la **media dei quadrati** dei valori della variabile meno il **quadrato della media**

# Misure di dispersione

## La Deviazione Standard

- **Problema:** la varianza non ha la stessa unità di misura dei dati
- Per questo motivo si utilizza la sua radice quadrata, ottenendo la **deviazione standard**:

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

- È chiamata anche scarto quadratico medio si può indicare con  $\sigma$  (o DS, o SQM)

# Misure di dispersione

## Il Coefficiente di Variazione

- Gli indici visti fino ad ora risentono dell'ordine di grandezza
- Difficile ipotizzare *confronti* tra variabili con unità di misura diversa

Si può ottenere un indicatore *puro*, dividendo la deviazione standard per la media aritmetica:

$$CV(x) = \frac{\sigma(x)}{|\bar{x}|} \cdot 100$$

### Commenti

- Possono esserci problemi di instabilità quando  $\bar{x} \approx 0$ .
- Devianza, Varianza, Deviazione Standard e CV sono legati alla somma dei quadrati degli scarti dalla media.
- Pertanto, sono sensibili alla presenza di **valori anomali**

# Misure di dispersione

## Esercizio 1

- Si dispone dei valori di emoglobina per un campione di  $n = 10$  lattanti
- $x_i = \{12, 2; 13, 2; 12, 3; 13, 1; 11, 5; 14, 5; 11, 3; 13, 6; 12, 6\}$
- Si calcolino: Varianza, Deviazione Standard, CV

# Misure di dispersione

## Esercizio 2

- Esami del primo anno sostenuti da due matricole

Voto	MAT.1 (x)	MAT.2(y)
Esame 1	23	18
Esame 2	24	21
Esame 3	24	24
Esame 4	24	24
Esame 5	24	27
Esame 6	25	30

Calcolare le seguenti quantità:

- Varianze:  $v^2(x) = \sigma_x^2$ ;  $v^2(y) = \sigma_y^2$
- Deviazioni Standard:  $v(x) = \sigma_x$ ;  $v(y) = \sigma_y$
- Coefficienti di Variazione:  $CV_x$ ;  $CV_y$

# Misure meno sensibili ai valori anomali

- Range (Intervallo di variabilità):

$$\max(x_i) - \min(x_i)$$

- Differenza Interquartile (InterQuartile Range)

$$IQR(x) = Q_3 - Q_1$$

- Altre misure di variabilità possono essere costruite attraverso:

$$\sum_i |x_i - \bar{x}| \quad \text{oppure} \quad \sum_i |x_i - \text{Med}(x)|$$

# Varianza “distorta” e “corretta”

Nella pratica, tutti i software riportano sempre le seguenti alternative.

- Varianza corretta:

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

- Deviazione standard corretta:  $s(x) = \sqrt{s^2(x)}$

La divisione per  $n - 1$  è legato ad alcune proprietà inferenziali (correttezza/non distorsione) che vedremo successivamente

Ovviamente:  $\hat{CV} = \frac{s(x)}{|\bar{x}|}$

# Utilizzare il foglio elettronico per calcolare la varianza

Come calcoliamo le misure di variabilità utilizzando il foglio elettronico (Excel) e le sue **funzioni**?

- Devianza: =DEV(DATI)
- Varianza: =VAR.C(DATI) (versione campionaria, corretta);  
=VAR.P(DATI) (versione di popolazione, distorta)
- Deviazione Standard: =DEV.ST.C(DATI) (versione campionaria, corretta); =DEV.ST.P(DATI) (versione di popolazione, distorta)
- Coefficiente di Variazione:  
 $\hat{CV} = DEV.ST.C(DATI) / MEDIA(DATI);$   
 $CV = DEV.ST.P(DATI) / MEDIA(DATI)$

# Riassunto: cosa abbiamo imparato oggi?

- ① Il concetto di variabilità
- ② Misure di dispersione: Devianza, Varianza e Deviazione Standard
- ③ Coefficiente di variazione
- ④ Altre misure di variabilità

# Materiale supplementare

## Esercizi per casa (1)

I  $n = 7$  ospedali di una città metropolitana hanno le seguenti disponibilità in termini di posti letto:

18; 49; 524; 67; 67; 53; 1370

- ① Calcolare la varianza del numero di posti letto
- ② Calcolare la deviazione standard del numero di posti letto
- ③ Calcolare il coefficiente di variazione (CV) del numero di posti letto
- ④ Calcolare il Range del numero di posti letto
- ⑤ Commentare i risultati ottenuti

# Materiale supplementare

## Esercizi per casa (2)

Si dispone di informazione sulla velocità massima di 16 calciatori.

<b>Giocatore</b>	G1	G2	G3	G4	G5	G6	G7	G8
<b>Speed (km/h)</b>	33,9	33,2	33,2	32,8	37,2	31,5	31,2	31,1
<b>Giocatore</b>	G9	G10	G11	G12	G13	G14	G15	G16
<b>Speed (km/h)</b>	31,0	30,3	29,8	29,7	29,3	25,6	20,5	20,3

- Calcolare la varianza della velocità massima
- Calcolare la deviazione standard della velocità massima
- Calcolare il Coefficiente di Variazione (CV) della velocità massima
- Calcolare il Range della velocità massima
- Commentare i risultati ottenuti

# Materiale supplementare

## Esercizi per casa (3, FACOLTATIVO; Livello “difficile” )

Alla fine dell’anno accademico si registra il numero di esami sostenuto da una coorte di 160 iscritte/i al primo anno di un corso di laurea.

N. Esami	Semplici			Cumulate		
	$n_i$	$f_i$	$p_i$	$N_i$	$F_i$	$p_i$
0	30					
1	23					
2	25					
3	18					
4	19					
5	20					
6	35					
<b>Tot.</b>				/	/	/

- Calcolare media e varianza
- Si riesce ad individuare qual è la mediana?

**Suggerimento:** media per tabelle di frequenze oppure usare il foglio elettronico

# Esercitazione

## Per casa

Calcolare le misure di variabilità (univariata) per il dataset “Pazienti” condiviso dal docente sulla classroom.

Alcune informazioni:

- FEV1: volume espiratorio forzato in 1 secondo. Il FEV1 in percentuale è ottenuto rapportando tale volume alla CVF (capacità vitale forzata), ossia massima quantità di aria che il paziente può espirare forzatamente dopo un'inspirazione massimale. Il FEV1 è il parametro di flusso utile nella diagnosi e nel monitoraggio dei pazienti affetti da patologie polmonari ostruttive (asma, broncopneumopatia cronica ostruttiva).
- La DLCO è il test che serve per esaminare lo scambio dei gas tra gli alveoli e i capillari polmonari (anche questo indicatore è espresso in %)
- BMI: indice di massa corporea = Peso / Altezza<sup>2</sup>

# Esercitazione

## Per casa

Calcolare le misure di variabilità (univariata) per il dataset “Pazienti” condiviso dal docente sulla classroom.

Alcune informazioni:

- Operative Time è il tempo in minuti dell'operazione chirurgica
- Hospital Stay è il tempo in giorni della degenza postoperatoria

**Domanda:** Quale degli indicatori utilizzati è più variabile?