

# Misure di variabilità: Varianza e deviazione standard

# Perché la variabilità?

**Tabelle e grafici sono ‘ingombranti’, lunghi da leggere e da riprodurre per analizzare le risposte di una collettività, perciò si sintetizzano le risposte con un valore medio, ad esempio la media.**

**Ma la sola media porta informazioni sufficienti per confrontare più gruppi di dati?  
Ecco un esempio per riflettere.**

# Scoprire la variabilità

Ecco i voti a un compito di matematica di ragazzi e ragazze

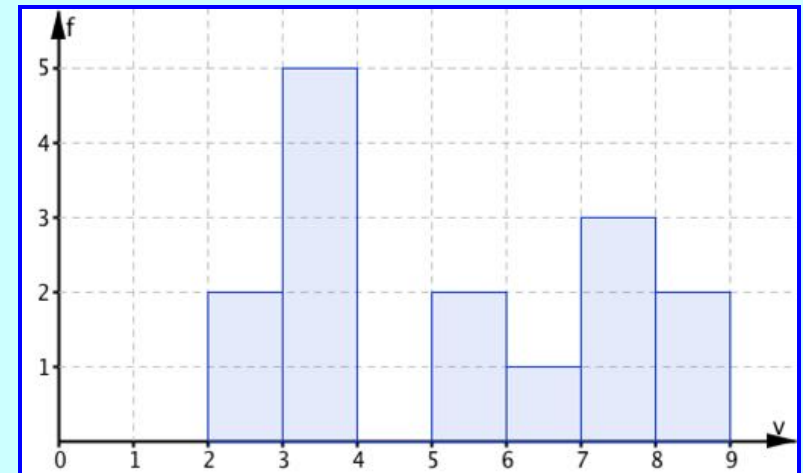
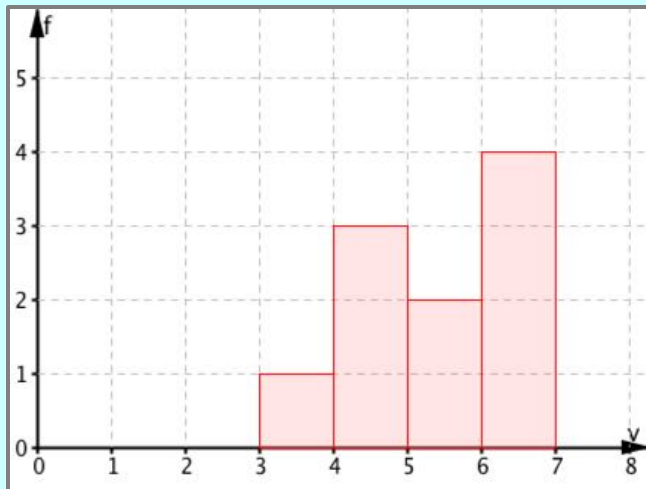
## Ragazze

Voto	Frequenza
4	1
5	3
6	2
6½	2
7	2
<b>Totale</b>	<b>10</b>
<b>Media</b>	<b>5,8</b>

## Ragazzi

Voto	Frequenza
3	2
4	5
6	2
7	1
8	3
9	2
<b>Totale</b>	<b>15</b>
<b>Media</b>	<b>5,8</b>

Stessa media



# Scoprire la variabilità

I voti a un compito di matematica di ragazzi e ragazze

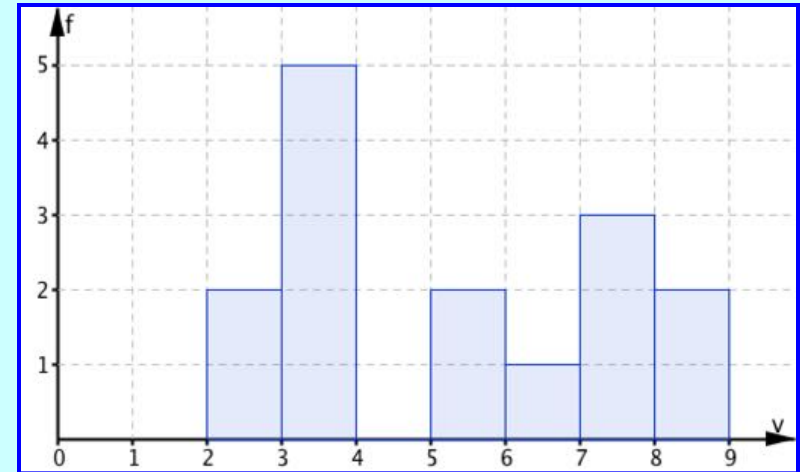
**Ragazze**



**I voti delle ragazze sono 'concentrati' vicino alla media**

**Le ragazze hanno voti molto simili, vicini alla media**

**Ragazzi**

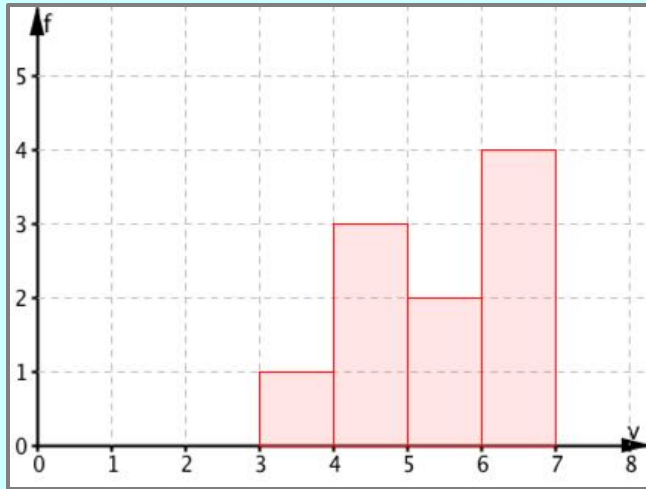


**I voti dei ragazzi sono 'dispersi' più lontano dalla media**

**Fra i ragazzi c'è un gruppo con voti insufficienti, compensato da un gruppo con voti molto buoni.**

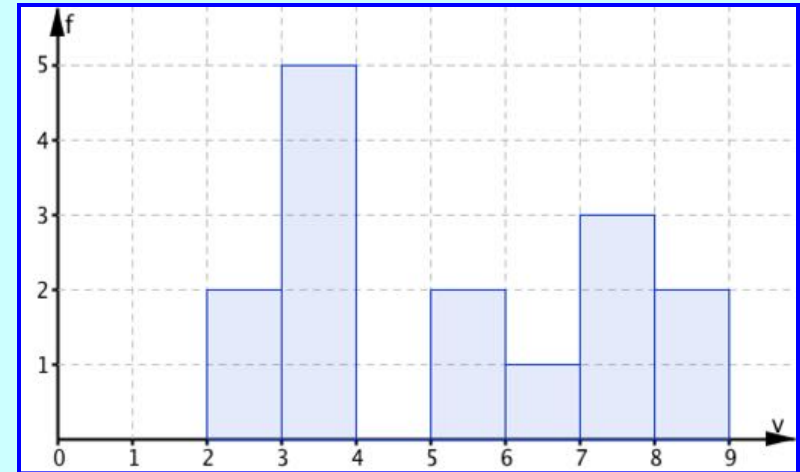
# Scoprire la variabilità

## Ragazze



I voti delle ragazze sono  
'concentrati' vicino alla media

## Ragazzi

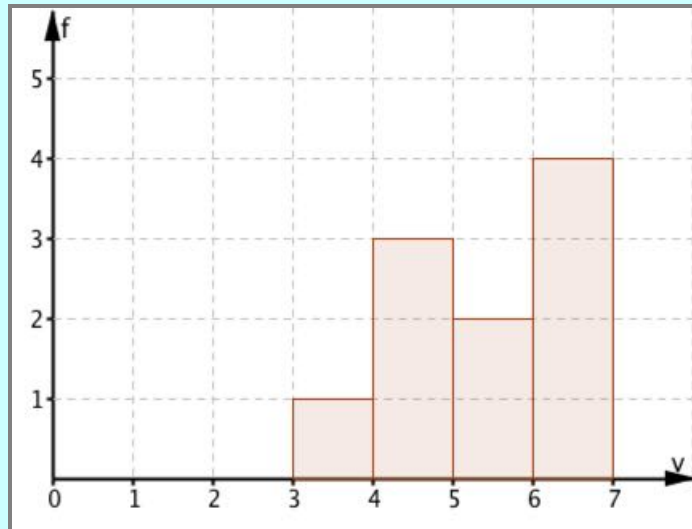


I voti dei ragazzi sono 'dispersi'  
più lontano dalla media

Il solo calcolo della media trascura  
*la dispersione o variabilità* dei dati

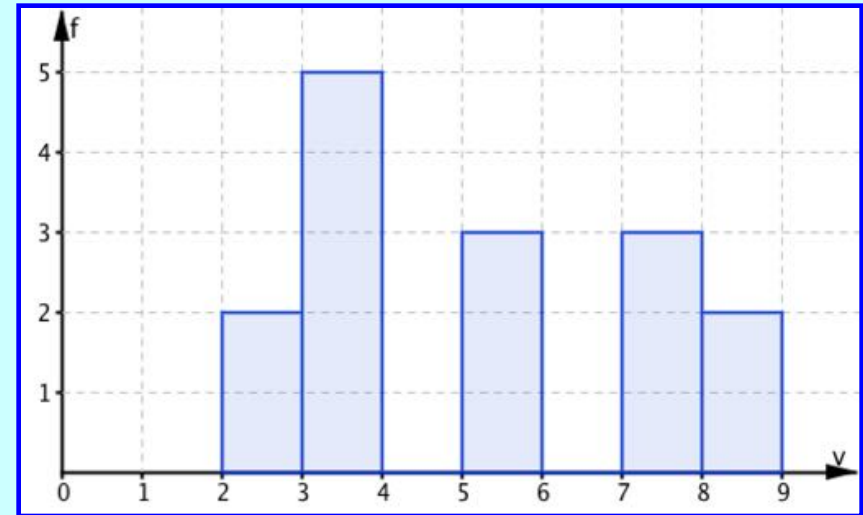
# Scoprire la variabilità

## Ragazze



I voti delle ragazze sono 'concentrati' vicino alla media

## Ragazzi



I voti dei ragazzi sono 'dispersi' più lontano dalla media

La valutazione della **variabilità** non può essere limitata all'aspetto dell'istogramma, perciò la statistica suggerisce vari metodi per esprimere la variabilità con un un numero.

# **Attività. Varianza e deviazione standard**

**Completa la scheda di lavoro per misurare la variabilità di un insieme di dati.**

# Che cosa hai trovato?

- **La varianza.**
- **La deviazione standard.**
- **Proprietà di varianza e deviazione standard.**

## Riflessioni sui risultati ottenuti



# A. La varianza

## Quesito1

Il gruppo delle 10 ragazze ha ottenuto i seguenti voti:

4 5 5 5 6 6 6½ 6½ 7 7

Il voto medio  $M_1$  del gruppo è 5,8.

Gli *scarti dalla media*, sono dati da:

$$4 - 5,8 = -1,8$$

$$5 - 5,8 = -0,8 \text{ (3 volte)}$$

$$6 - 5,8 = 0,2 \text{ (2 volte)}$$

$$6,5 - 5,8 = 0,7 \text{ (2 volte)}$$

$$7 - 5,8 = 1,2 \text{ (2 volte)}$$

a. Quanto vale la somma  $S$  degli scarti?  $S = 0$

b. Che cosa puoi dire sul segno degli scarti?

Gli scarti positivi sono compensati dagli scarti negativi.

c. Perché non puoi valutare la variabilità con la media  $M_S$  degli scarti?

Perché otterrei sempre  $M_S = 0$ .

d. Per superare le difficoltà legate al segno degli scarti, la statistica suggerisce di valutare la variabilità con la *varianza*  $\sigma^2$ , data dalla media dei quadrati degli scarti.

$$\sigma^2 = \frac{(4 - 5,8)^2 + (5 - 5,8)^2 \cdot 3 + (6 - 5,8)^2 \cdot 2 + (6,5 - 5,8)^2 \cdot 2 + (7 - 5,8)^2 \cdot 2}{10} = 0,91$$

$$M_s = \frac{S}{10} = \frac{0}{10} = 0$$

# B. Varianza e deviazione standard

## Quesito 2

Per avere una valutazione della variabilità confrontabile con i dati, la statistica introduce la *deviazione standard*  $\sigma$ , detta anche ‘scarto quadratico medio’.

Deviazione standard dei voti delle ragazze:

$$\sigma = \sqrt{\sigma^2} \cong 0,95$$

# B. Varianza e deviazione standard

## Quesito 3: due serie di dati 'insoliti'

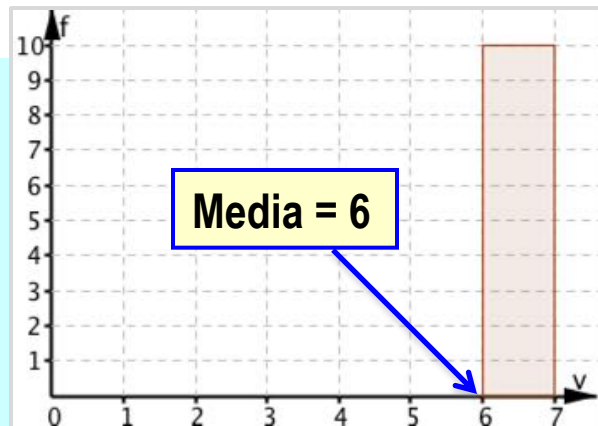
A partire dalle due serie di dati rappresentati dalle tabelle valuta:

Dati **A**, media  $M_A = 6$ , varianza  $\sigma_A^2 = 0$ , deviazione standard  $\sigma_A = 0$

Dati **B**, media  $M_B = 6$ , varianza  $\sigma_B^2 = 16$ , deviazione standard  $\sigma_B = 4$

**A**

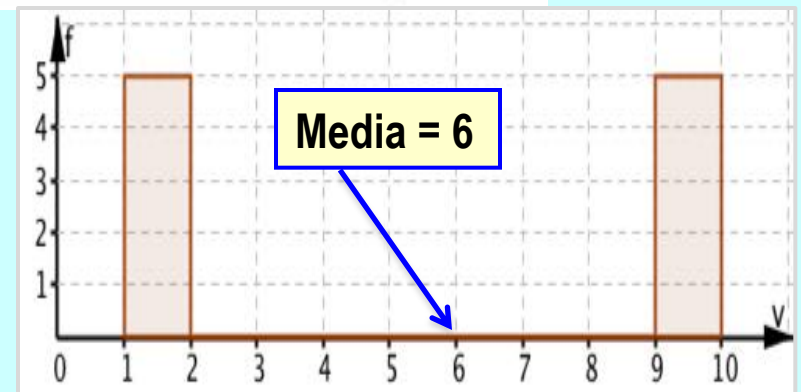
Voto	Frequenza
6	10



Dati tutti uguali, ottengo  
 $\sigma^2 = 0$  e quindi  $\sigma = 0$

**B**

Voto	Frequenza
2	5
10	5



Dati lontani dalla media, ottengo  
 $\sigma^2 = 16$  e quindi  $\sigma = 4$

# Vocabolario statistico

## Indici statistici

**Varianza e deviazione standard sono anche detti '*Indici di variabilità*' o '*Indici di dispersione*'.**

**E così si dice che media, mediana e moda sono '*Indici di posizione*' o '*Indici di tendenza centrale*'**

# C. Varianza e deviazione standard con foglio di calcolo

## Quesito 4

Esegui il lavoro richiesto e scrivi qui sotto le risposte alle domande.

a. Completa la tabella qui sotto

<b>Voti ragazze</b>	Media = 5,8	Varianza $\approx$ 0,91	Dev. Stand. $\approx$ 0,95
<b>Voti ragazzi</b>	Media = 5,8	Varianza $\approx$ 4,56	Dev. Stand. $\approx$ 2,14

b. Cosa osservi se confronti la varianza di ragazze e ragazzi?

La varianza dei ragazzi è più grande (circa quattro volte)

c. Cosa osservi se confronti la dev. standard di ragazze e ragazzi?

La dev. standard dei ragazzi è più grande (circa il doppio)

	A	B	C	D
1	RAGAZZE	RAGAZZE	RAGAZZI	RAGAZZI
2	4		3	
3	5		3	
4	5		4	
5	5		4	
6	6		4	
7	6		4	
8	6.5		4	
9	6.5		6	
10	7		6	
11	7		7	
12	Media	5.8	8	
13	Varianza	0.91	8	
14	Dev.Stand.	0.95	8	
15			9	
16			9	
17			Media	5.8
18			Varianza	4.56
19			Dev.Stand.	2.14

### Con il foglio di calcolo

**Media:** `Media[C2:C16]`

**Varianza:** `Varianza[C2:C16]`

**Deviazione standard:** `DS[C2:C16]`

# D. Media, varianza e deviazione standard con un dato 'anomalo'

## Quesito 5

Qual è l'effetto dell'ultimo dato tanto più grande degli altri?

Sostituire solo 7 con 70, molto più grande degli altri dati, ha fatto passare:

- la **media** da 5,8 a 12,1;
- la **varianza** da 0,91 a 373,24;
- la **deviazione standard** da 0,95 a 19,32.

Media, varianza e deviazione standard sono fortemente influenzate dai dati anomali.

RAGAZZE	RAGAZZE	VOTO NUOVO	VOTO NUOVO
4		4	
5		5	
5		5	
5		5	
6		6	
6		6	
6.5		6.5	
6.5		6.5	
7		7	
7		70	
<b>Media</b>	<b>5.8</b>	<b>Media2</b>	<b>12.1</b>
<b>Varianza</b>	<b>0.91</b>	<b>Varianza2</b>	<b>373.24</b>
<b>Dev.Stand.</b>	<b>0.95</b>	<b>Dev.Stand2</b>	<b>19.32</b>

# E. Proprietà di varianza e deviazione standard

## Quesito 6

RAGAZZE	RAGAZZE	RAGAZZE*2	RAGAZZE*2
4		8	
5		10	
5		10	
5		10	
6		12	
6		12	
6.5		13	
6.5		13	
7		14	
7		14	
Media	5.8	Media2	11.6
Varianza	0.91	Varianza2	3.64
Dev.Stand.	0.954	Dev.Stand2	1.908

$$3,64 = 0,91 \times 4$$

$$1,908 = 0,954 \times 2$$

- a. Se moltiplico per 2 tutti i dati,  
la varianza  $\sigma^2$  è moltiplicata per 4  
la deviazione standard  $\sigma$  è moltiplicata per 2

# E. Proprietà di varianza e deviazione standard

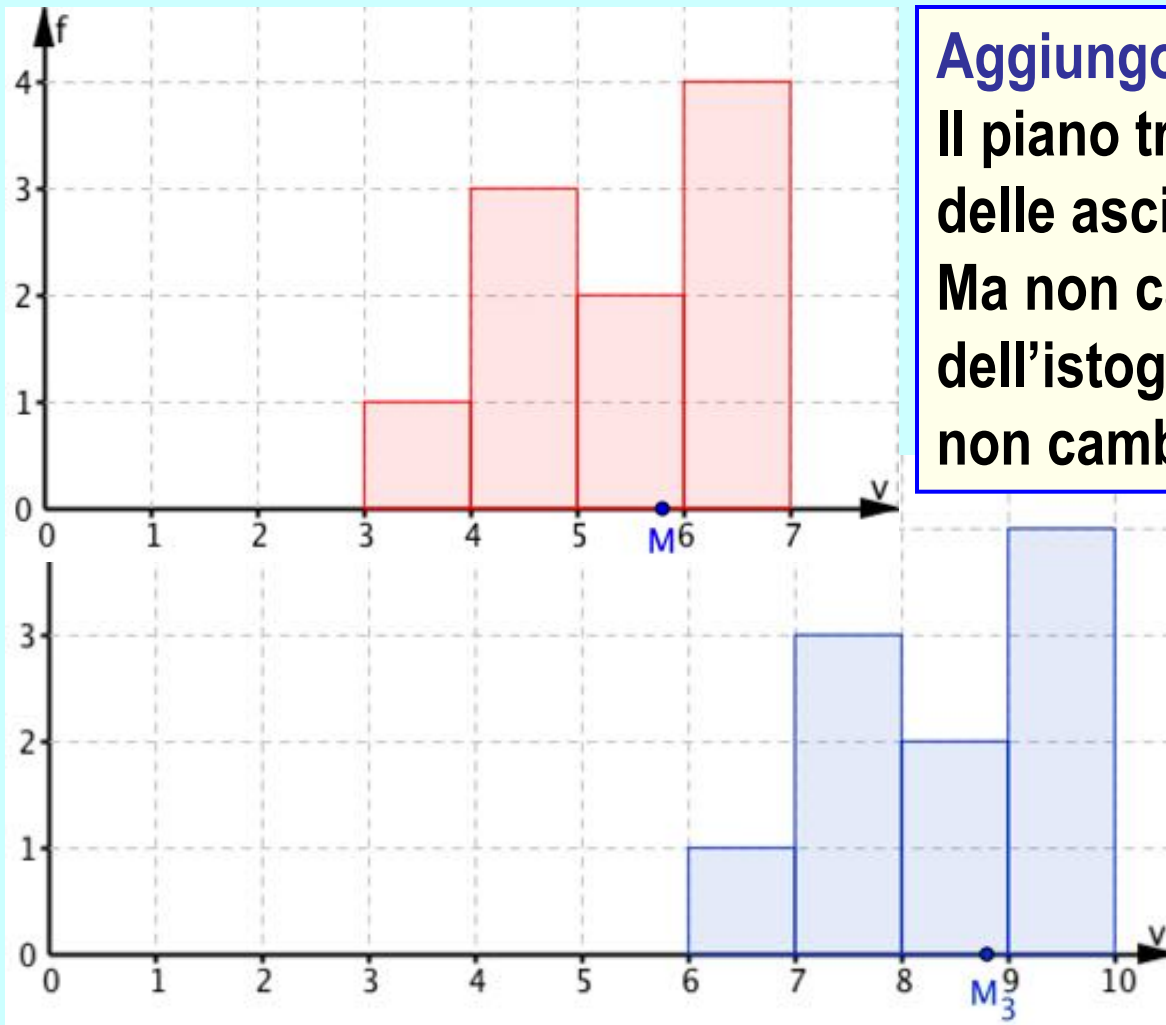
## Quesito 6

RAGAZZE	RAGAZZE	RAGAZZE+3	RAGAZZE+3
4		7	
5		8	
5		8	
5		8	
6		9	
6		9	
6.5		9.5	
6.5		9.5	
7		10	
7		10	
<b>Media</b>	<b>5.8</b>	<b>Media3</b>	<b>8.8</b>
<b>Varianza</b>	<b>0.91</b>	<b>Varianza3</b>	<b>0.91</b>
<b>Dev.Stand.</b>	<b>0.954</b>	<b>Dev.Stand3</b>	<b>0.954</b>

b. Se aggiungo 3 a tutti dati,  
la varianza  $\sigma^2$  e la deviazione standard  $\sigma$  non cambiano

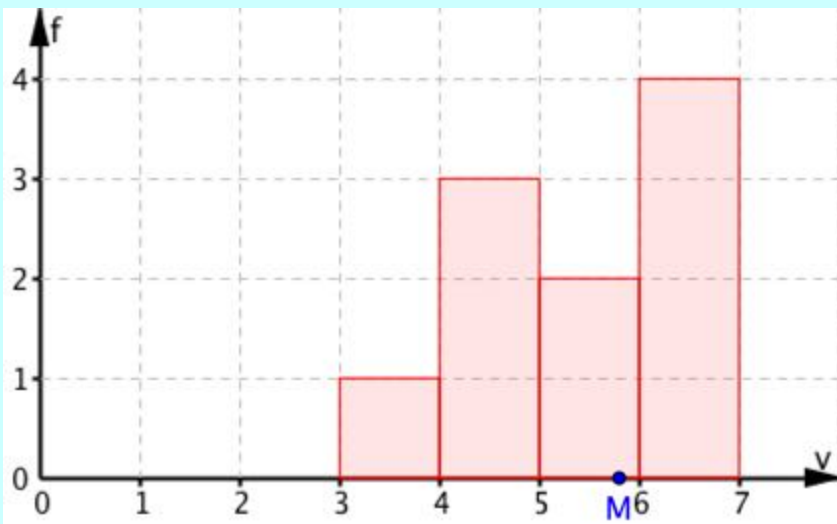


# Proprietà di varianza e deviazione standard: un'interpretazione grafica

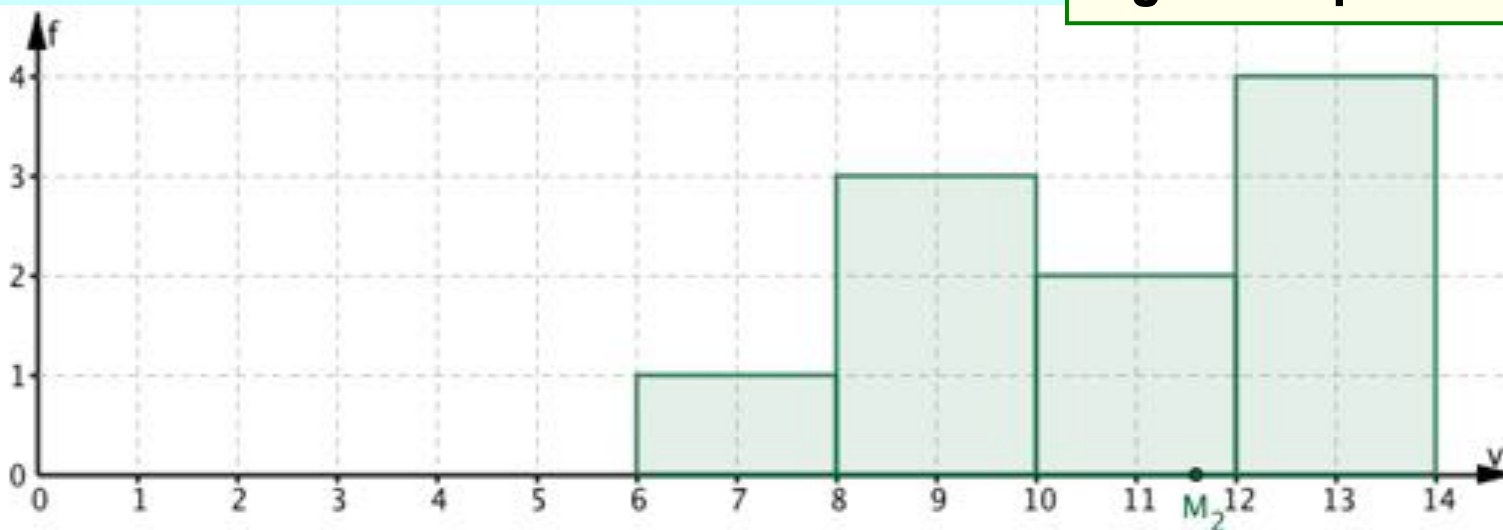


Aggiungo 3 a tutti i dati.  
Il piano trasla lungo l'asse delle ascisse e 'trascina'  $M$ .  
Ma non cambia la forma dell'istogramma. Perciò non cambiano  $\sigma$  e  $\sigma^2$ .

# Proprietà di varianza e deviazione standard: un'interpretazione grafica



**Moltiplico per 2 tutti i dati.**  
**Dilato il piano lungo l'asse delle ascisse e raddoppio 'la larghezza' dell'istogramma.**  
**Perciò raddoppia  $\sigma$  legato ai dati, ma quadruplica  $\sigma^2$  legato al quadrato dei dati.**



# F. Proprietà di varianza e deviazione standard: validità generale

## Quesito 7

Indico con  $a, b, c$  tre dati, con  $M$  la loro media, con  $\sigma^2$  la loro varianza e con  $\sigma$  la loro deviazione standard. Risulta:

$$M = \frac{a+b+c}{3} \quad \sigma^2 = \frac{(a-M)^2 + (b-M)^2 + (c-M)^2}{3} \quad \sigma = \sqrt{\sigma^2}$$

a. Se moltiplico per  $k$  tutti i dati, ottengo:

$$M_1 = \frac{ka + kb + kc}{3} = kM$$

$$\sigma_1^2 = \frac{(ka - kM)^2 + (kb - kM)^2 + (kc - kM)^2}{3} = \frac{k^2 \left[ (a-M)^2 + (b-M)^2 + (c-M)^2 \right]}{3} = k^2 \sigma^2$$

$$\sigma_1 = \sqrt{\sigma_1^2} = k\sigma$$

**Se moltiplico per  $k$  tutti i dati**, anche la media  $M$  e la deviazione standard  $\sigma$  sono moltiplicate per  $k$ ; invece la varianza  $\sigma^2$  è moltiplicata per  $k^2$ .

# F. Proprietà di varianza e deviazione standard: validità generale

## Quesito 7

Indico con  $a, b, c$  tre dati, con  $M$  la loro media, con  $\sigma^2$  la loro varianza e con  $\sigma$  la loro deviazione standard. Risulta:

$$M = \frac{a+b+c}{3} \quad \sigma^2 = \frac{(a-M)^2 + (b-M)^2 + (c-M)^2}{3} \quad \sigma = \sqrt{\sigma^2}$$

b. Se aggiungo  $h$  a tutti i dati, ottengo:

$$M_2 = \frac{a+h+b+h+c+h}{3} = \frac{a+b+c+3h}{3} = M+h$$

$$\sigma_2^2 = \frac{[a+h-(M+h)]^2 + [b+h-(M+h)]^2 + [c+h-(M+h)]^2}{3} = \frac{(a-M)^2 + (b-M)^2 + (c-M)^2}{3} = \sigma^2$$

$$\sigma_2 = \sigma$$

**Se aggiungo  $h$  a tutti i dati**, anche alla media  $M$  aggiungo  $h$ ; invece la varianza  $\sigma^2$  e deviazione standard  $\sigma$  non cambiano.

# G. Un procedimento alternativo per calcolare la varianza

## Quesito 8

Indico con  $a$ ,  $b$ ,  $c$  tre dati; completa il procedimento qui sotto calcolare la varianza.

- Calcolo la media  $M = \frac{a+b+c}{3} \Leftrightarrow 3M = a+b+c$

- Calcolo la varianza

$$\begin{aligned}\sigma^2 &= \frac{(a-M)^2 + (b-M)^2 + (c-M)^2}{3} = \frac{a^2 + b^2 + c^2 - 2M(a+b+c) + 3M^2}{3} = \\ &= \frac{a^2 + b^2 + c^2 - 6M^2 + 3M^2}{3}\end{aligned}$$

- Concludo che posso calcolare la varianza anche con la formula

$$\sigma^2 = \frac{a^2 + b^2 + c^2}{3} - M^2$$

# Simboli per scrivere formule generali

**Per avere una formula generale, adatta a calcolare varianza e deviazione standard, ho indicato 3 dati qualunque con le lettere  $a$ ,  $b$ ,  $c$ .**

**Ma così la formula non è del tutto generale: i dati sono solo 3, mentre in statistica i dati possono essere 100, 1000, ... non basterebbero tutte le lettere dell'alfabeto per indicarli.**

# Simboli per scrivere formule generali

Per questo si indica:

- il numero di dati con  $N$ ;
- i singoli dati con  $x_1, x_2, \dots, x_N$

Attenzione a scrittura e lettura dei simboli:

Simbolo	Si legge	Significa
$x^2$	<i>x alla seconda</i>	$x^2 = x \cdot x$
$x_2$	<i>x con due</i>	Secondo dato
$x^1$	<i>x alla prima</i>	$x^1 = x$
$x_1$	<i>x con uno</i>	Primo dato
$x_1^2$	<i>x con uno al quadrato</i>	$x_1^2 = x_1 \cdot x_1$

## Il simbolo di 'sommatoria'

Per calcolare, ad esempio, la media di 100 dati, prima di tutto debbo addizionare i 100 numeri ed è molto lunga da scrivere questa addizione. Per abbreviare la scrittura si usa il simbolo di 'sommatoria'  $\Sigma$

Due esempi.

$$x_1 + x_2 + x_3 + x_4 = \sum_{i=1}^{i=4} x_i$$

$$a_1 + a_2 + a_3 + a_4 + a_5 = \sum_{k=1}^{k=5} a_k$$



# Formule statistiche generali

**Con questi simboli puoi scrivere formule generali per calcolare gli indici statistici che hai studiato finora.**

# Media, varianza e deviazione standard

I dati sono elencati uno di seguito all'altro

Indico  $N$  dati con i simboli  $x_1, x_2, \dots, x_N$

- La **media**  $M$  è data da:

$$M = \frac{x_1 + x_2 + \dots + x_N}{N} = \sum_{k=1}^{k=N} \frac{x_k}{N}$$

- La **varianza**  $\sigma^2$  è data da:

$$\sigma^2 = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_N - M)^2}{N} = \sum_{k=1}^{k=N} \frac{(x_k - M)^2}{N}$$

oppure

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_N^2}{N} - M^2 = \sum_{k=1}^{k=N} \frac{x_k^2}{N} - M^2$$

- La **deviazione standard**  $\sigma$  è data da:

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

# Media, varianza e deviazione standard

I dati sono raccolti in una tabella di frequenza.

## ESEMPIO

Voto	5	6	7	Totale	Media
Frequenza	3	4	3	10	6

- La **media**  $\bar{x}$  è data da:

$$\bar{x} = \frac{5 \times 3 + 6 \times 4 + 7 \times 3}{10} = 6$$

- La **varianza**  $\sigma^2$  è data da:

$$\sigma^2 = \frac{(5-6)^2 \times 3 + (6-6)^2 \times 4 + (7-6)^2 \times 3}{10} = 0,6$$

oppure

$$\sigma^2 = \frac{5^2 \times 3 + 6^2 \times 4 + 7^2 \times 3}{10} - 6^2$$

- La **deviazione standard**  $\sigma$  è data da:

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

## IN GENERALE

Dato	$x_1$	...	$x_p$	Totale	Media
$f$	$f_1$	...	$f_p$	$N$	$M$

- La **media**  $M$  è data da:

$$M = \frac{x_1 \cdot f_1 + \dots + x_p \cdot f_p}{N} = \sum_{k=1}^{k=p} \frac{x_k \cdot f_k}{N}$$

- La **varianza**  $\sigma^2$  è data da:

$$\sigma^2 = \frac{(x_1 - M)^2 \cdot f_1 + \dots + (x_p - M)^2 \cdot f_p}{N} = \sum_{k=1}^{k=p} \frac{(x_k - M)^2}{N}$$

oppure

$$\sigma^2 = \frac{x_1^2 \cdot f_1 + \dots + x_p^2 \cdot f_p}{N} - M^2 = \sum_{k=1}^{k=p} \frac{x_k^2 \cdot f_k}{N} - M^2$$

- La **deviazione standard**  $\sigma$  è data da:

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