

DOVE ERAVAMO RINASTI....

- TESTIAMO la VARIABLE interesse dell'indagine statistica su un CAMPIONE di DIMENSIONE N.

- TROVIAMO I VALORI $y_1 \quad y_2 \quad \dots \quad y_N$
 $\downarrow$ $\downarrow$
sulla I UNITÀ sulla N-esima UNITÀ
STATISTICA STATISTICA
 $y_i \equiv$ valore trovato sulla
i-esima unità statistica

- SE SI RIPETONO, CONVIENE PRENDERE SOLI QUELLI DISTINTI TRA LORO.

LI CHIAMIAMO $z_1 \dots z_k$ con $k \leq N$

- "CONTIAMO" QUANTE VOLTE CIASCUNO z_i COMPARE SUL CAMPIONE. COSTRUIAMO LA TABELLA

MODALITÀ	$z_1 \dots z_k$	DISTINTE
FREQUENZE ASSOLUTE	$f_1 \dots f_k$	

///
numero di volte in cui
ogni modalità compare
sul campione. Quindi:
 $\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k = N$

- Poiché le f_i possono assumere valori molto grandi, conviene lavorare con le **FREQUENZE RELATIVE**

$$p_i := \frac{f_i}{N} \leq 1$$

$$\frac{f_i}{N} \leq \frac{N}{N} = 1$$

Allora $0 < p_i \leq 1$

$$\begin{aligned} p_1 + p_2 + \dots + p_N &= \\ &= \frac{f_1}{N} + \frac{f_2}{N} + \dots + \frac{f_N}{N} = \frac{\sum_{i=1}^N f_i}{N} = \frac{N}{N} \\ &= 1 \end{aligned}$$

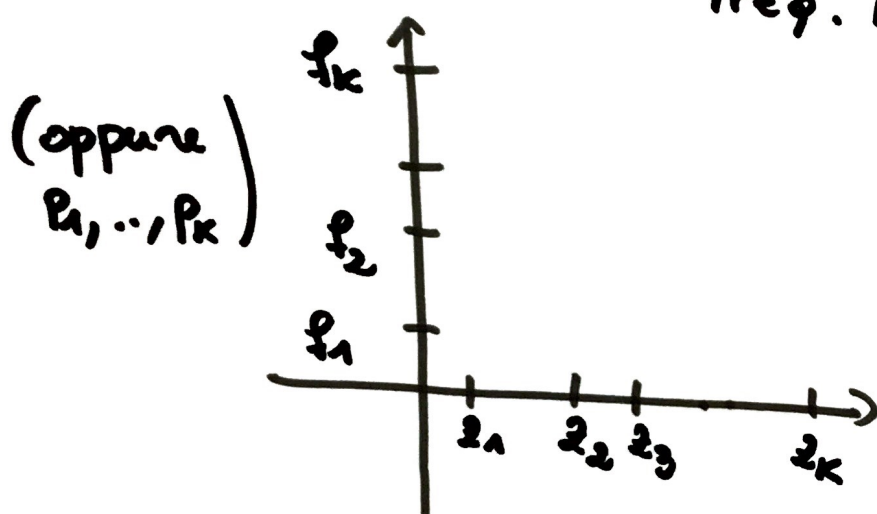
- Poi abbiamo introdotto la **moda** $\equiv$ modalità che più compare sul campione

- Se i dati sono tanti oppure ~~sono~~ se la variabile è continua, conviene raccoglierci in classi del tipo
incluso $\rightarrow [x_i, x_{i+1}) \leftarrow$ escluso

Allora $f_i \equiv$ somma di tutti le volte in cui compaiono dati compresi tra x_i e x_{i+1} (escluso)

RAPPRESENTAZIONE

- PIANO CARTESIANO
- sull'asse $x \rightarrow$ MODALITÀ $z_1, \dots, z_k$
- sull'asse $y \rightarrow$ Freq. ASS. f_i oppure Freq. REL. p_i



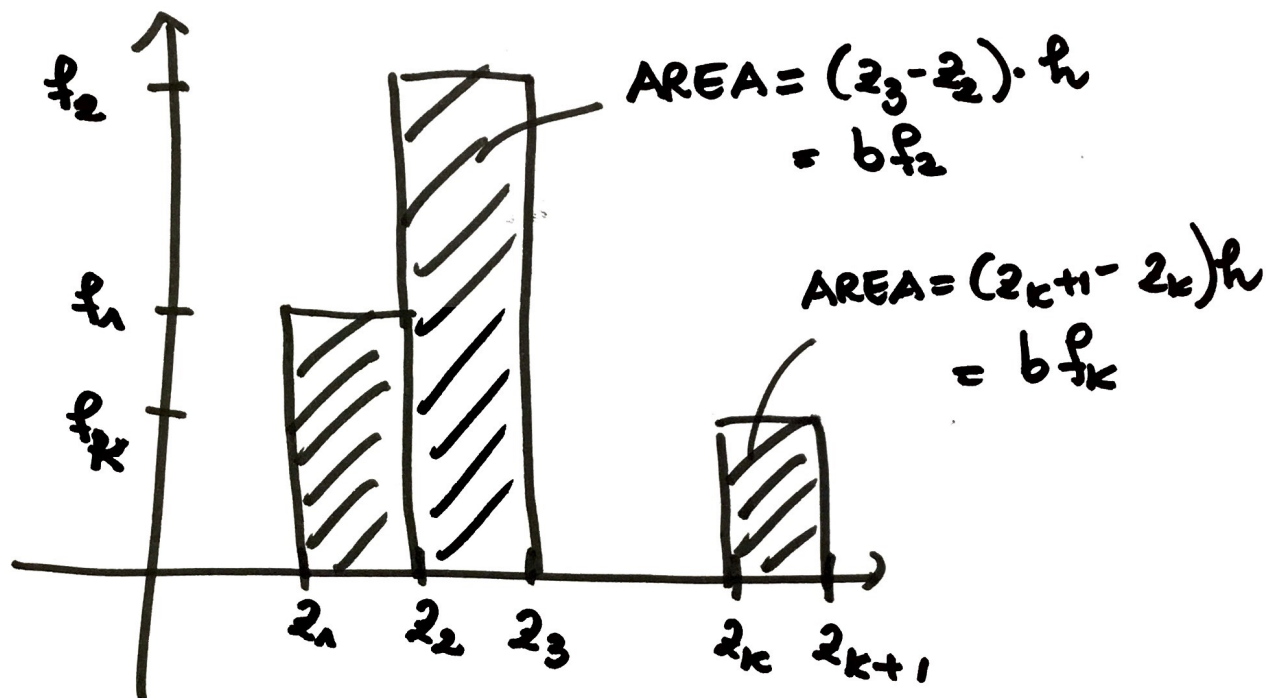
- traccio bastoncini di altezza uguale (o proporzionale) a f_i (o p_i) $\hookrightarrow$ A BASTONCINI
- unisco i punti di coordinate (z_i, f_i) con una poligonale $\hookrightarrow$ A POLIGONALI
- se dò spessore ai bastoncini $\hookrightarrow$ A BARRE O COLONNE

Se i dati sono raccolti in classi :
ISTOGRAMMI

I CASO

Tutte le classi hanno la
stessa ampiezza

$$z_{i+1} - z_i = b$$



Tutte le aree dei rettangoli
costruiti sono proporzionali.

(con fattore di proporzionalità b)
alle frequenze assolute

(stesso discorso con le freq. relative)

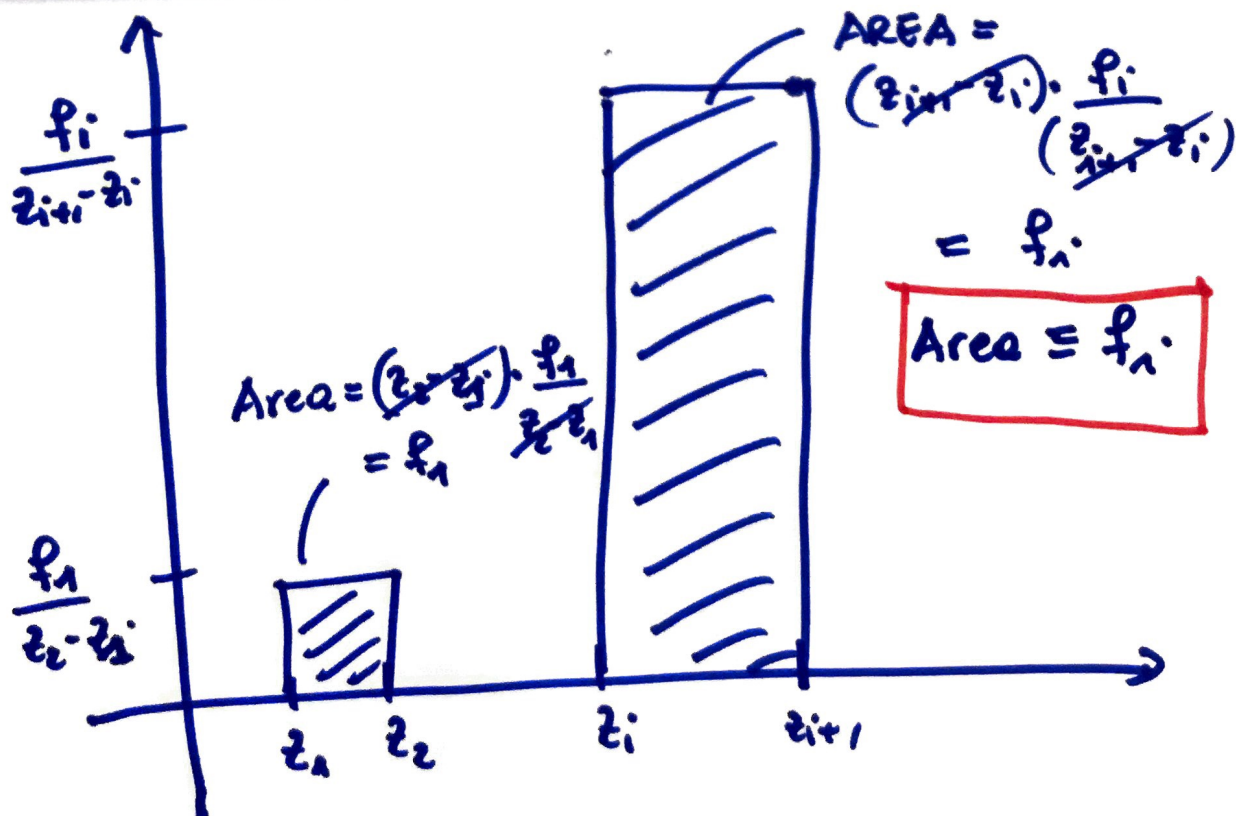
II CASO : ANPIEZZE DELLE CLASSI DIVERSE
 Allora le altezze dei rettangoli sono
 scelte in modo che le Aree siano
 uguali alle frequenze.

Se le altezze sono $\frac{f_i}{z_{i+1} - z_i}$ allora

l'area $\bar{x} = \text{base} \cdot \text{altezza} =$

$$= \underbrace{\frac{f_i}{z_{i+1} - z_i}}_{\text{altezza}} \cdot \underbrace{(z_{i+1} - z_i)}_{\text{base}} = f_i$$

$$\Rightarrow \text{Area} = f_i$$



CON IL SIMBOLO

$$\sum_{i=1}^N x_i \text{ intendo la}$$

Somma di $x_1, x_2, \dots, x_N$.

Si legge "Sommatore
per i che va da 1 a N "

Infatti:

Faccio
la
somma
di tutti

$$\sum_{i=1}^N x_i$$

indica l'ultimo termine
della somma x_N

intende il 1° termine
della somma x_1

MEDIA: cerco $\bar{y}$ tale che

$$\underbrace{(\bar{y} - y_1)}_{\text{scostamento da } y_1} + (\bar{y} - y_2) + \dots + (\bar{y} - y_N) = 0$$

$\underbrace{\hspace{10em}}_{\text{scostamento da } y_N}$

SOMMA DEGLI "SCOSTAMENTI" = 0

$$\Rightarrow N\bar{y} = y_1 + \dots + y_N$$

$$\Rightarrow \bar{y} = \frac{y_1 + y_2 + \dots + y_N}{N} = \frac{\sum_{i=1}^N y_i}{N}$$

← MEDIA

Questa formula viene usata quando ho tutti gli N dati "alla rinfusa"

Se invece li ho organizzati con la tabella

	z_1	$\dots$	z_k
Freq. Ass.	f_1	$\dots$	f_k

per calcolare la media dovrai sommare

$$\underbrace{z_1 + \dots + z_1}_{f_1 \text{ volte}} + \underbrace{z_2 + \dots + z_2}_{f_2 \text{ volte}} + \dots$$

$$= f_1 \cdot z_1 + f_2 \cdot z_2 + \dots + f_k \cdot z_k$$

La formula diventa

$$\begin{aligned}\bar{y} &= \frac{z_1 \cdot f_1 + z_2 \cdot f_2 + \dots + z_k \cdot f_k}{N} \\ &= \frac{\sum_{i=1}^k z_i \cdot f_i}{N} \\ &= \frac{\sum_{i=1}^k z_i \cdot f_i}{\sum_{i=1}^k f_i}\end{aligned}$$

Se ho le frequenze relative

$$\begin{array}{c|c} z_1 & \dots & z_k \\ \hline p_1 & \dots & p_k \end{array}$$

per calcolare $\bar{y} = \frac{\sum_{i=1}^k f_i \cdot z_i}{N}$
uso il fatto che

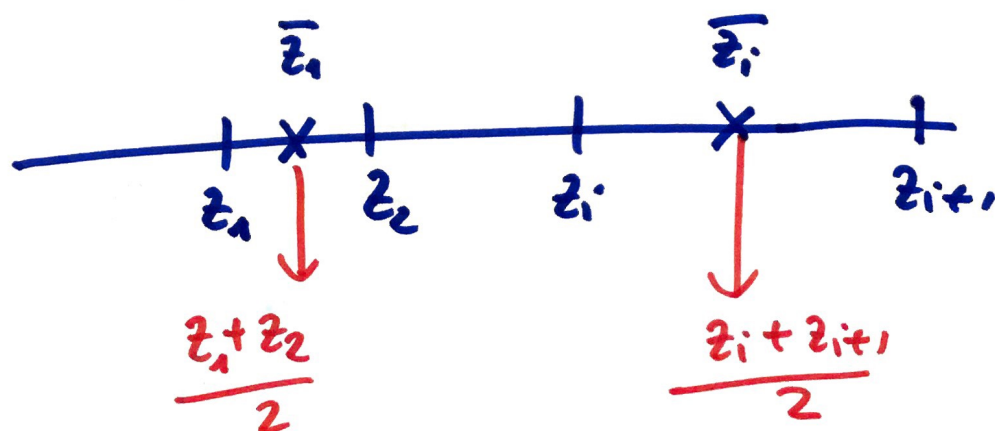
$$\begin{aligned}p_i &= \frac{f_i}{N} \quad \text{ossia} \quad f_i = p_i \cdot N \\ \Rightarrow \bar{y} &= \frac{\sum_{i=1}^k f_i \cdot z_i}{N} = \frac{\sum_{i=1}^k N p_i \cdot z_i}{N} = \cancel{N} \frac{\sum_{i=1}^k p_i \cdot z_i}{\cancel{N}} \\ &= \sum_{i=1}^k p_i \cdot z_i\end{aligned}$$

NON È PIÙ
NECESSARIO N

Se i dati sono organizzati in classi:

	$\bar{z}_1 = \frac{z_1 + z_2}{2}$	$\bar{z}_2 = \frac{z_2 + z_3}{2}$	
	(z_1, z_2)	(z_2, z_3)	$\dots$
FREQ	f_1	f_2	
ASS			

sostituisco ad ogni classe (z_i, z_{i+1})
il punto medio tra z_i e z_{i+1}



ottengo la tabella

	$\bar{z}_1$	$\bar{z}_2 \dots$	$\bar{z}_i$
	f_1	f_2	f_i

$$\Rightarrow \bar{y} = \frac{f_1 \cdot \bar{z}_1 + f_2 \cdot \bar{z}_2 + \dots}{N}$$

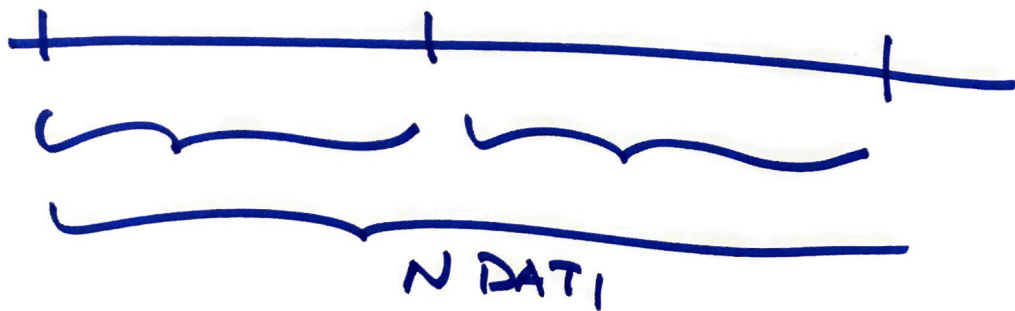
MEDIANA

ORDINO TUTTI i dati in modo crescente
Per semplicità, gi  suppongo

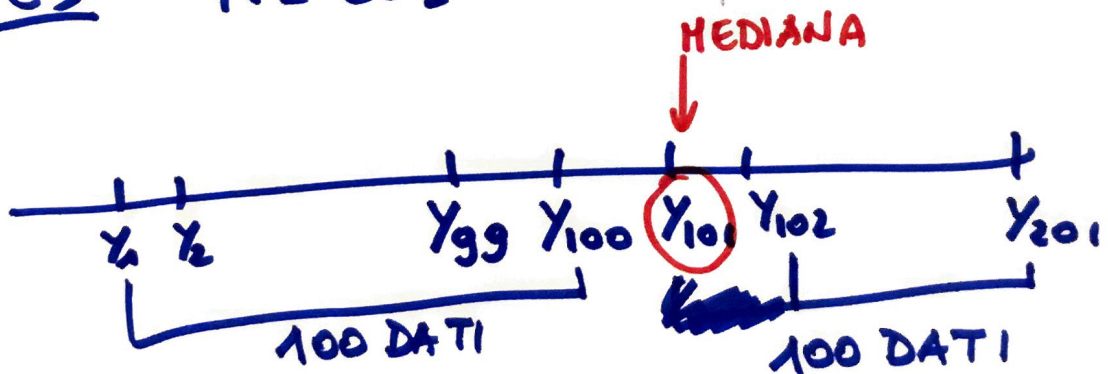
$$Y_1 \leq Y_2 \dots \leq Y_N$$

La mediana   il dato che sta
nel mezzo di questi dati
ordinati.

MEDIANA



ES $N = 201$



Se N   PARI, prendo i dati "centrali"
e faccio la loro media