

Corso di Statistica Psicometrica

Canale 1 - Matricole Dispari

CdS triennale in
Scienze e Tecniche
Psicologiche

Domenico Vistocco

Dipartimento di Scienze Politiche
Università degli Studi di Napoli Federico II



Attenzione...



Lei ha il diritto di rimanere in silenzio: qualsiasi cosa dirà / farà potrà essere usata contro di lei...

Nella
puntata
precedente...



- Tipologia di variabili
- Tipologia di tabelle
- Frequenza assolute, relative e percentuali
- Frequenze cumulate
- Densità di frequenza
- Rappresentazioni grafiche
- Indici di posizione: la classe particolare degli indici di tendenza centrale
- Proprietà di media aritmetica e mediana
- Modelli statistici: i centri come “semplici” modelli
- Misura della variabilità
- Indici di disuguaglianza
- Indici di dispersione
- Indici basati sul confronto tra posizione opposte
- Un indicatore robusto di variabilità: il MAD (median absolute deviation)
- I coefficienti di variazione
- Proprietà desiderabili di un indice di variabilità e comportamento delle tre classi di indicatori
- Tre particolari trasformazioni lineari: centratura, riduzione e standardizzazione
- **Proprietà di decomposizione della devianza / varianza**

Dati divisi in gruppi: sintesi (2)

Gruppo	Unità	Cardinalità	Medie	Devianze
1	$x_{1,1}, \dots, x_{1,j}, \dots, x_{1,n_1}$	n_1	$\bar{x}_1$	Dev ₁
2	$x_{2,1}, \dots, x_{2,j}, \dots, x_{2,n_2}$	n_2	$\bar{x}_2$	Dev ₂
⋮	⋮	⋮		
i	$x_{i,1}, \dots, x_{i,j}, \dots, x_{i,n_i}$	n_i	$\bar{x}_i$	Dev _i = $\sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$
⋮	⋮	⋮		
G	$x_{G,1}, \dots, x_{G,j}, \dots, x_{G,n_G}$	n_G	$\bar{x}_G$	Dev _G

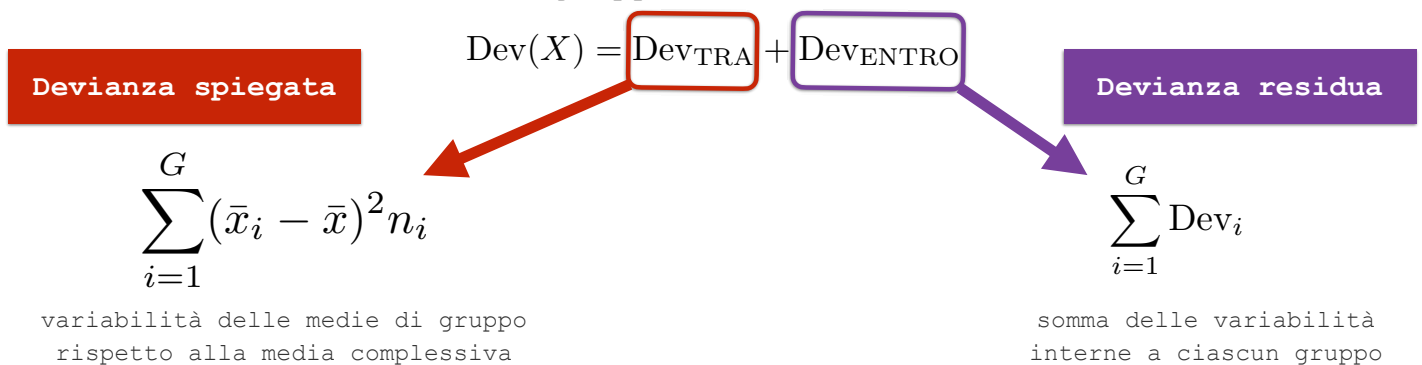
$$n = \sum_{i=1}^G n_i$$

Proprietà di decomposizione della devianza

La devianza complessiva:

$$\text{Dev}(X) = \sum_{i=1}^G \sum_{j=1}^{n_j} (x_{i,j} - \bar{x})^2$$

Può essere scomposta in devianza esterna (tra i gruppi o among) e devianza interna (entro i gruppi o within):



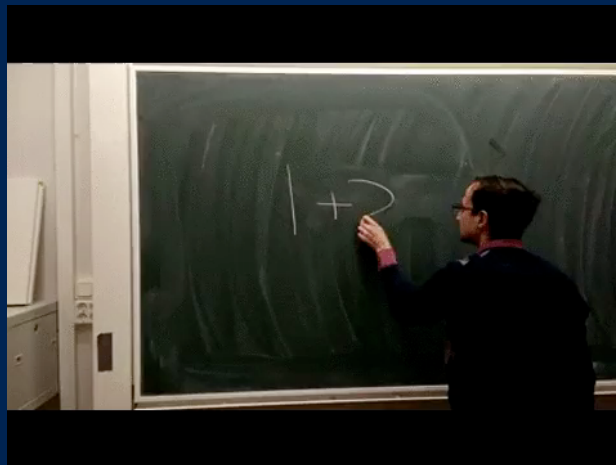
Proprietà di decomposizione della devianza

The diagram shows two formulas. On the left, the F-statistic is defined as $F = \frac{\text{Devianza spiegata} / (G - 1)}{\text{Devianza residua} / (n - G)}$. On the right, the eta-squared coefficient is defined as $\eta^2 = \frac{\text{Devianza spiegata}}{\underbrace{\text{Devianza residua} + \text{Devianza spiegata}}_{\text{Devianza totale}}}$.

Formula alternativa di calcolo della varianza / devianza

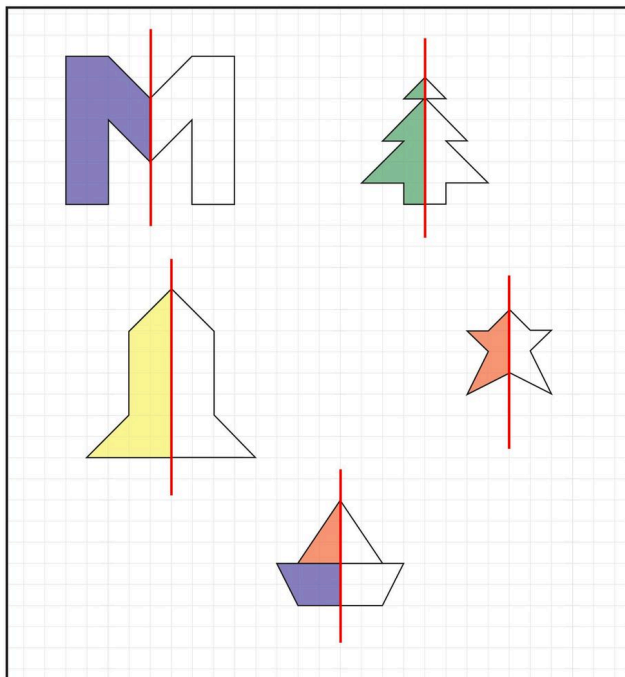
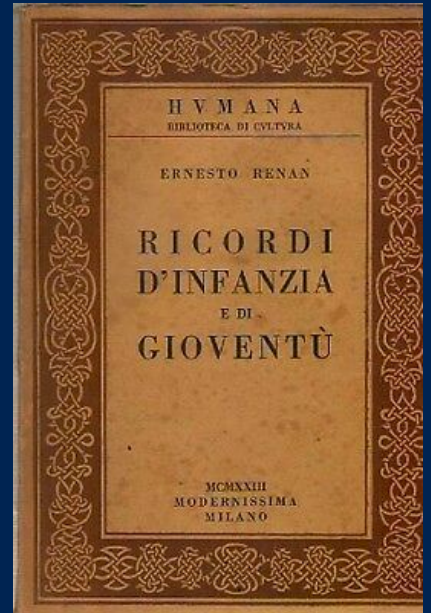
varianza →
$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

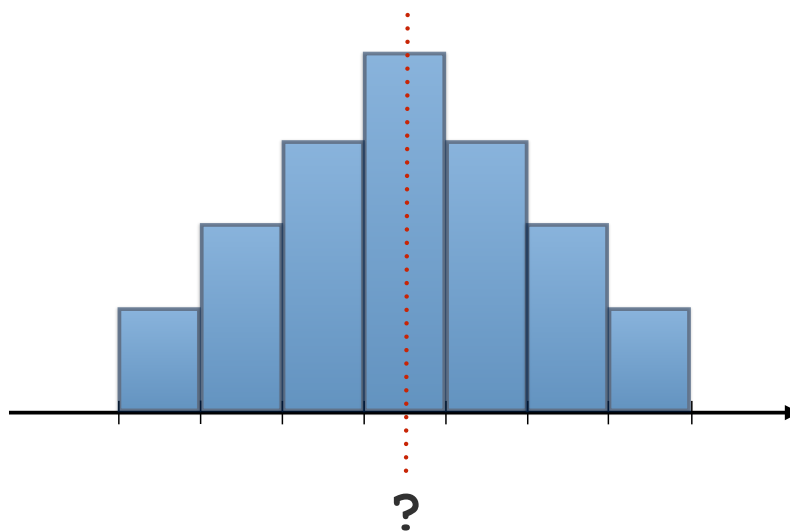
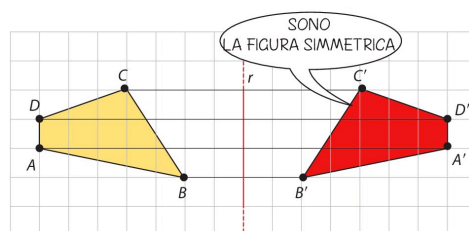
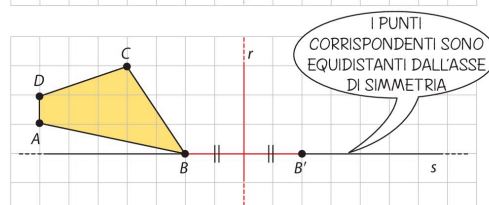
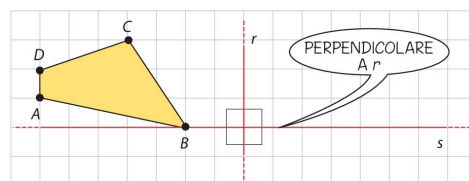
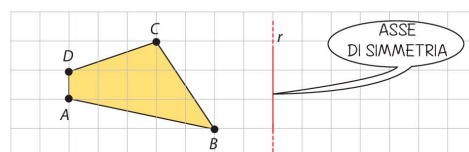
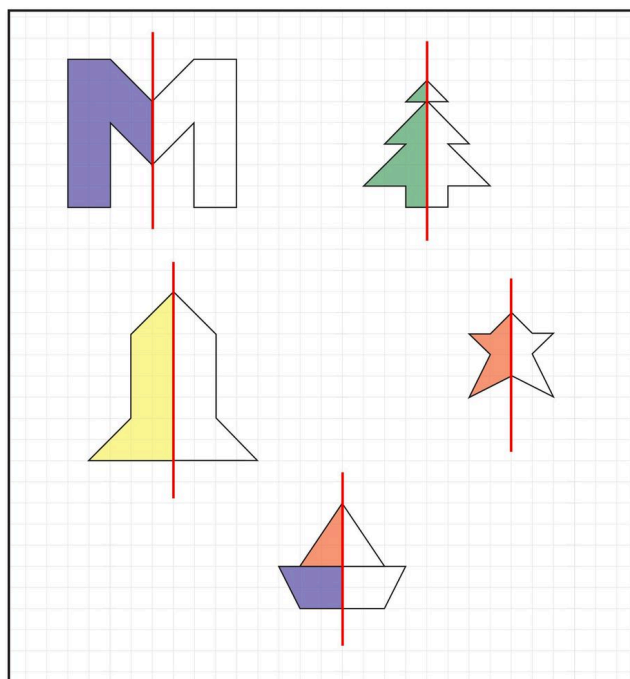
devianza →
$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n \bar{x}^2$$

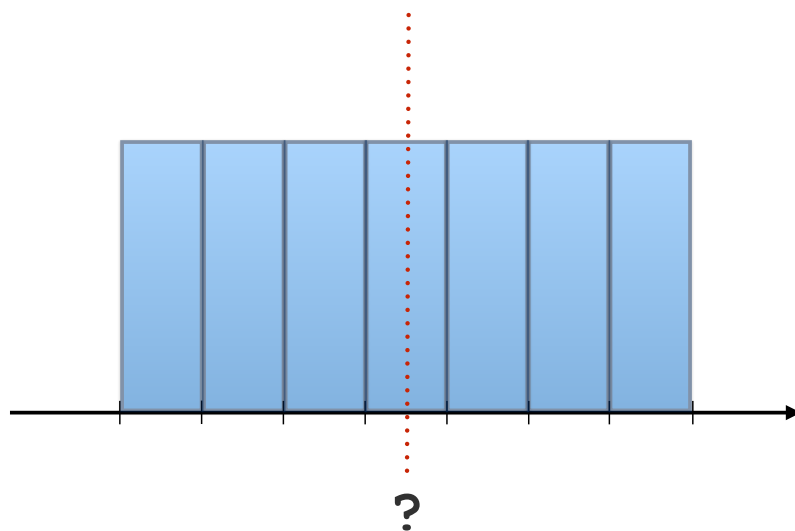
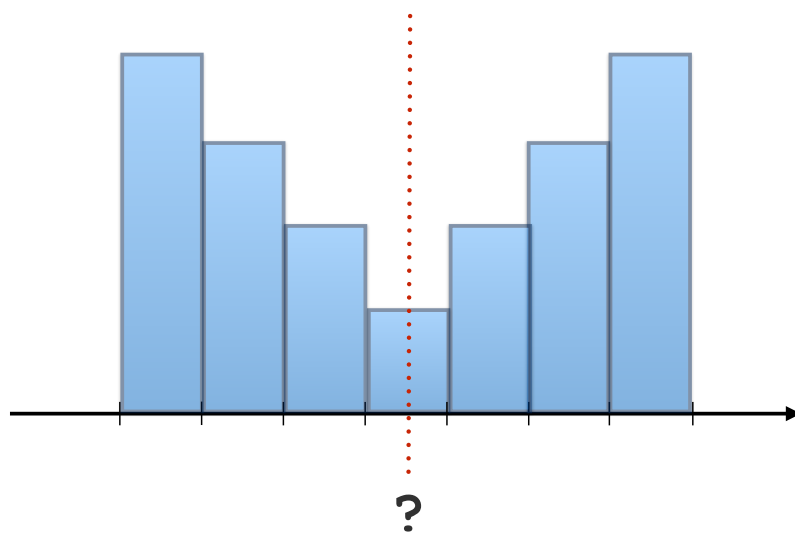


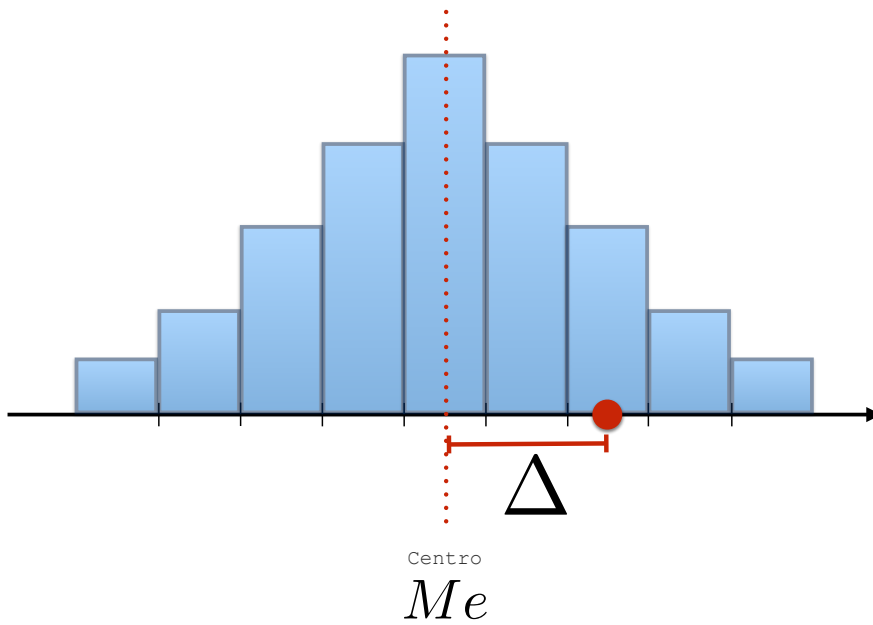
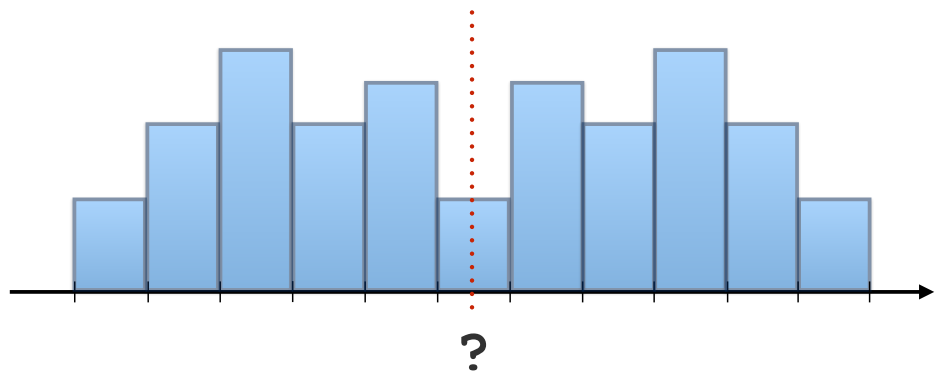
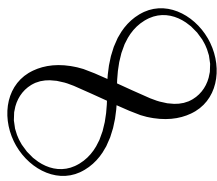
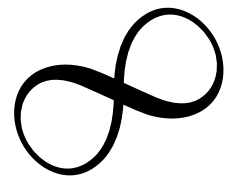
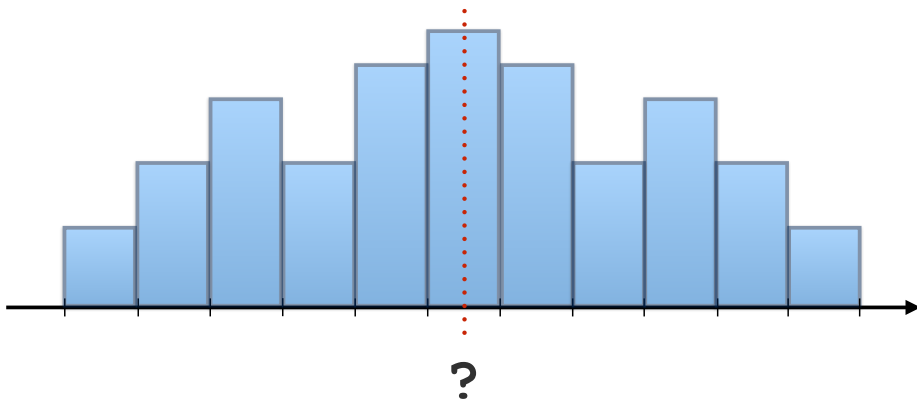
Si riparte?

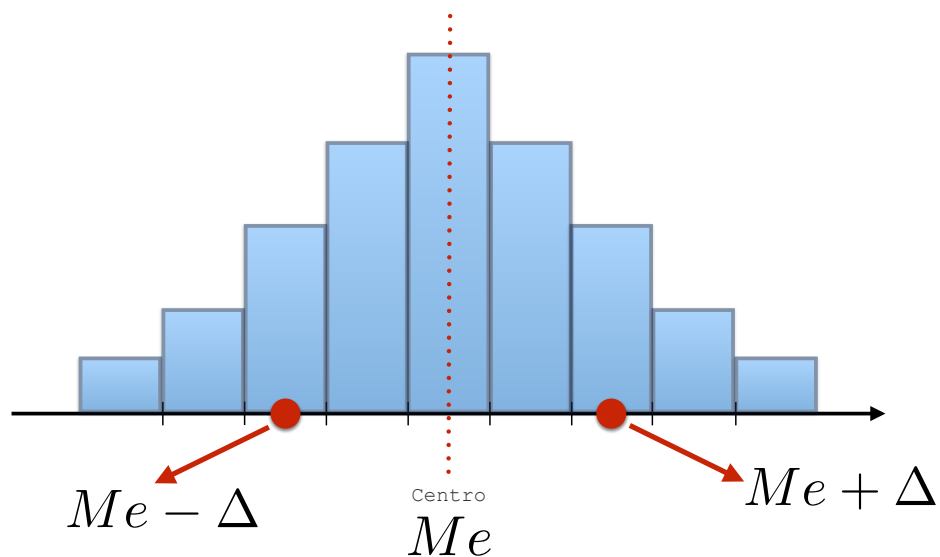
Amarcord











Il ruolo degli scarti



Il ruolo degli scarti...

distribuzione
simmetrica

Ad ogni scarto negativo corrisponderà uno
scarto positivo di uguale grandezza

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

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



Il ruolo degli scarti...

distribuzione
simmetrica

Ad ogni scarto negativo corrisponderà uno
scarto positivo di uguale grandezza

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

$$\frac{\sum_{i=1}^n (x_i - \bar{x})}{n} = 0$$



Il ruolo degli scarti...

distribuzione
simmetrica

Ad ogni scarto negativo corrisponderà uno
scarto positivo di uguale grandezza

$$\sum_{i=1}^n (x_i - \bar{x})^3$$



$$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}$$

Il ruolo degli scarti: l'indice di Fisher

distribuzione
asimmetrica

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n} \neq 0$$

Asimmetria negativa
(sinistra)

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n} < 0$$

la "maggior" parte dei
valori è concentrato sulla
destra e c'è una coda più
pesante sul lato sinistro

Asimmetria positiva
(destra)

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n} > 0$$

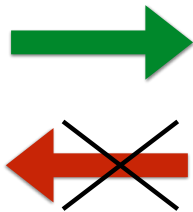
la "maggior" parte dei
valori è concentrato sulla
sinistra e c'è una coda più
pesante sul lato destro

Attenzione ai falsi segnali



Il ruolo degli scarti: l'indice di Fisher

Distribuzione
simmetrica



$$\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n} = 0$$

NOTA: questo vale per tutti gli indici di forma (possono dare segnali imprecisi)

Dettagli tecnici...

Il ruolo degli scarti: l'indice di Fisher

unità di misura originaria

$$\sqrt[3]{\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}}$$

indice relativo

$$\begin{aligned} \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}}{s^3} &= \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n s^3} = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})^3}{s}}{n} \\ &\parallel \\ &\frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3}{n} \\ &\parallel \\ \boxed{\gamma_3} &= \sum_{i=1}^n \frac{z_i^3}{n} \end{aligned}$$

Il ruolo degli scarti: l'indice di Fisher

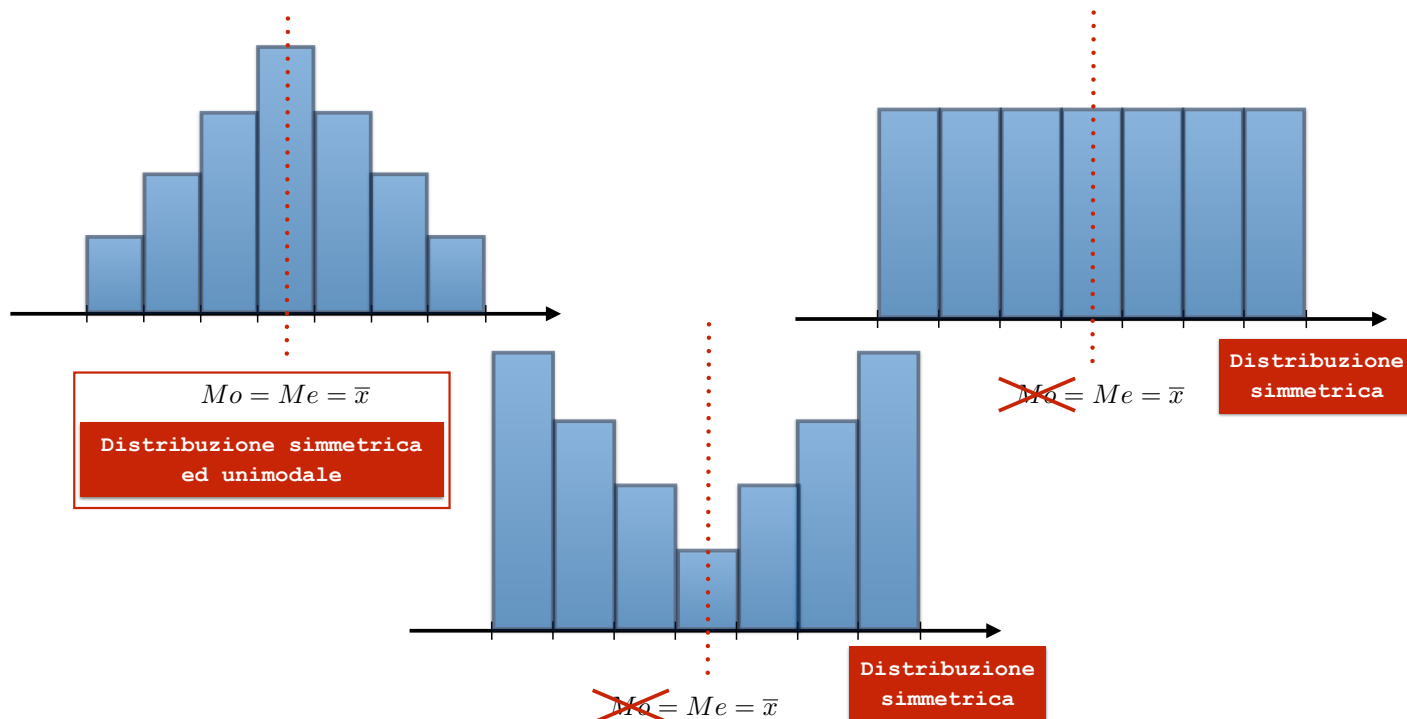
The diagram illustrates the derivation of the Fisher index formula through three stages, each represented by a blue arrow pointing to a mathematical expression:

- Serie grezza** (Raw data) points to the formula:
$$\frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}^3}{s} \right)}{n}$$
- Tabella di frequenza per singole modalità** (Frequency table for single modalities) points to the formula:
$$\frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}^3}{s} \right) n_i}{n}$$
- Tabella di frequenza per classi di modalità** (Frequency table for classes of modalities) points to the formula:
$$\frac{\sum_{i=1}^n \left(\frac{c_i - \bar{x}^3}{s} \right) n_i}{n}$$

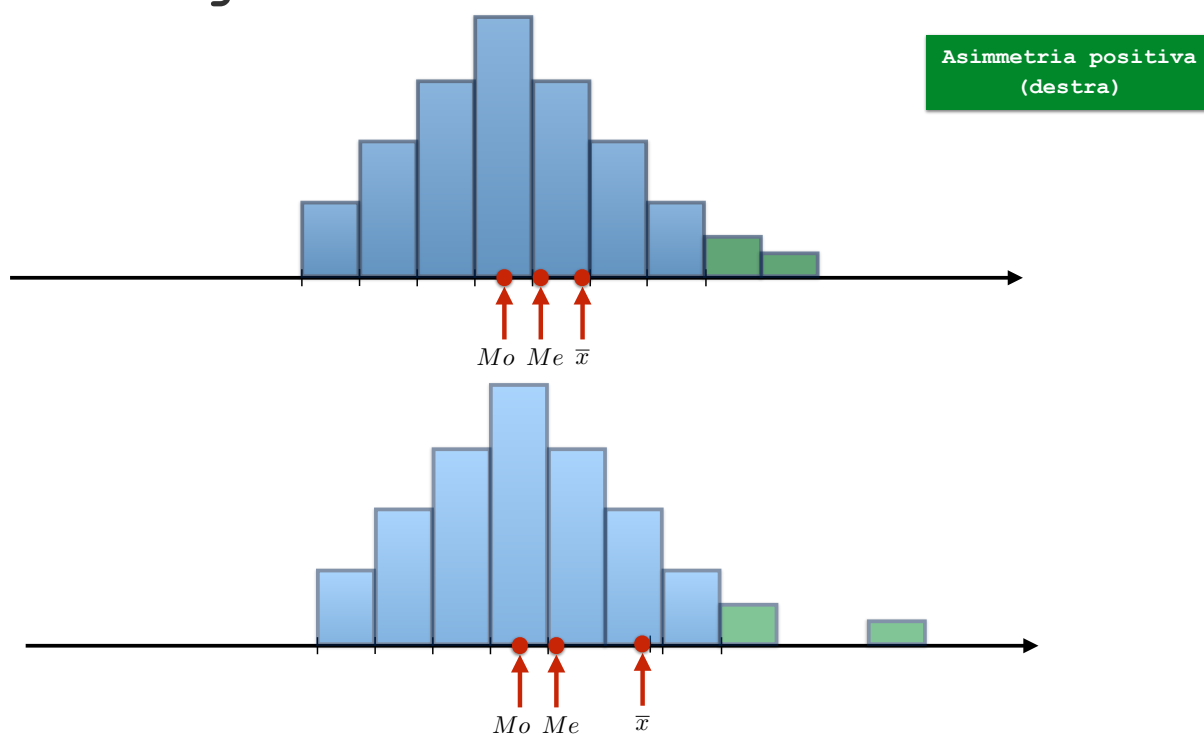
A red dotted arrow points from the c_i term in the third formula to the definition: $c_i = \hat{x}_i = \frac{x_{i-1} + x_i}{2}$

**Altre "forme" per
la forma: il ruolo
dei centri**

Il ruolo degli indici di tendenza centrale

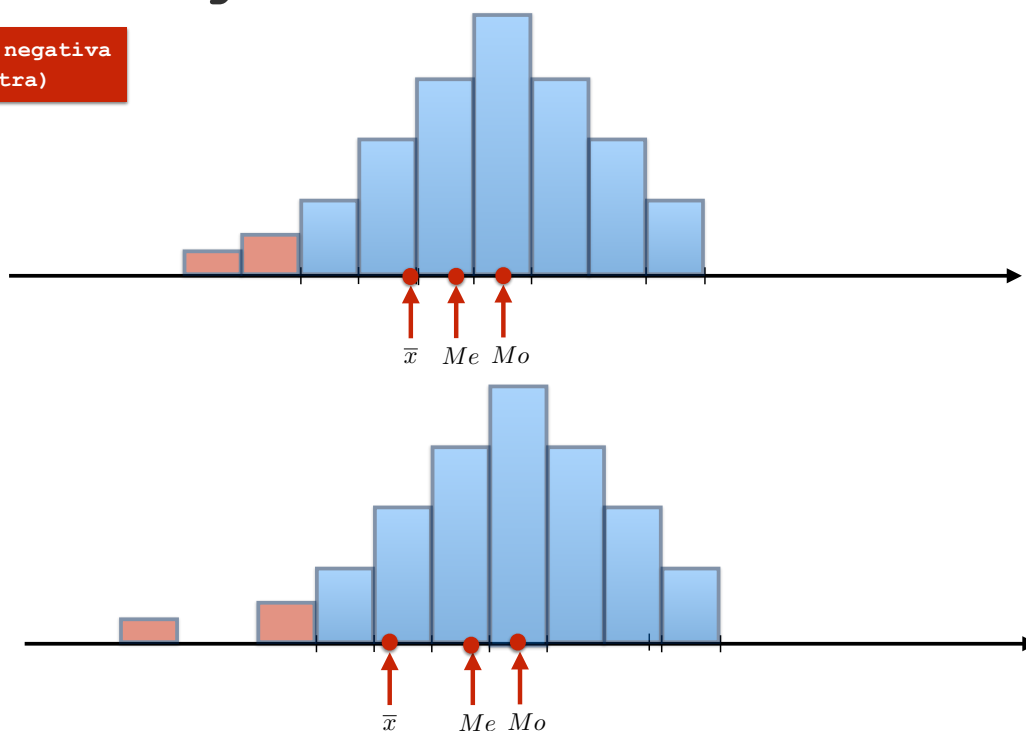


Il ruolo degli indici di tendenza centrale

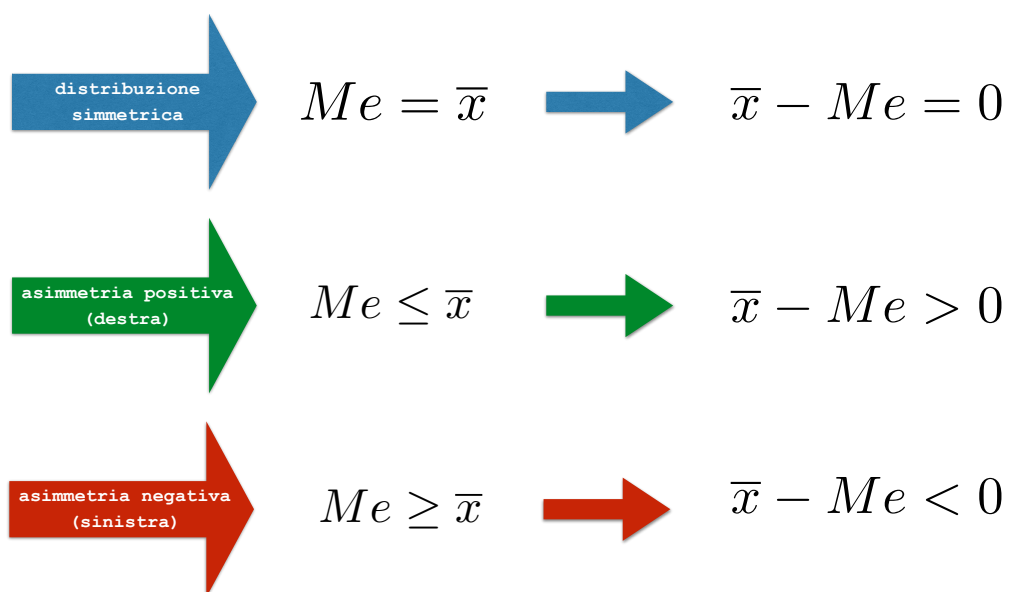


Il ruolo degli indici di tendenza centrale

Asimmetria negativa
(sinistra)



Il ruolo degli indici di tendenza centrale: l'indice di Hotelling - Solomon



Il ruolo degli indici di tendenza centrale: l'indice di Hotelling - Solomon



$$\frac{\bar{x} - Me}{s}$$

Hotelling e Solomon hanno dimostrato che:

$$|\bar{x} - Me| \leq s$$



$$-1 \leq \frac{\bar{x} - Me}{s} \leq +1$$

Attenzione ai falsi segnali



Il ruolo degli indici di tendenza centrale: l'indice di Hotelling - Solomon



2, 8, 10, 10, 10, 15, 15

$$Mo = Me = \bar{x}$$



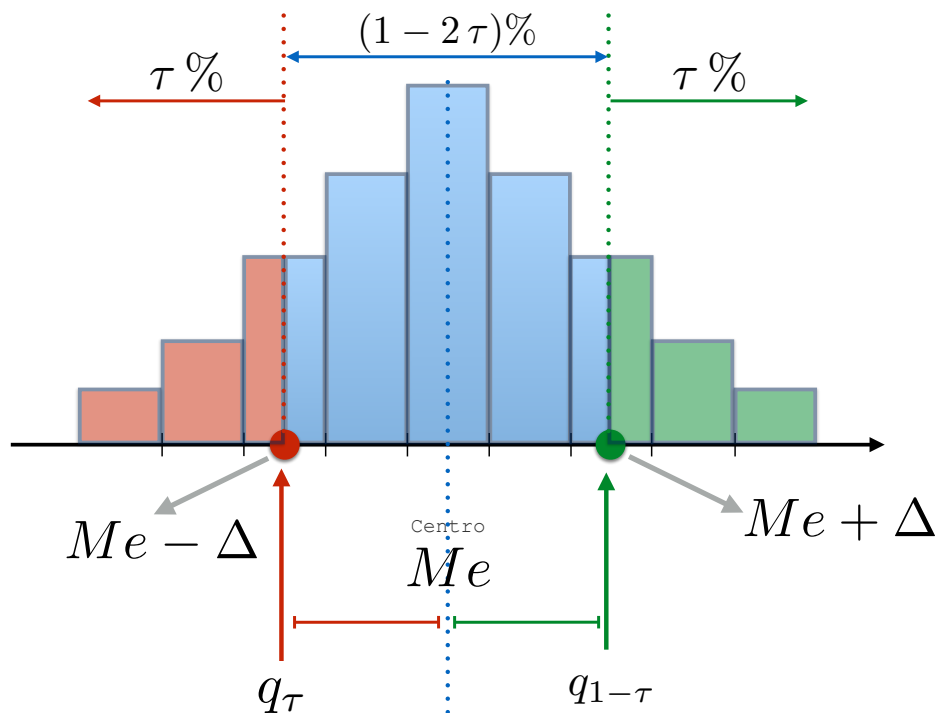
-8, -2, 0, 0, 0, +5, +5



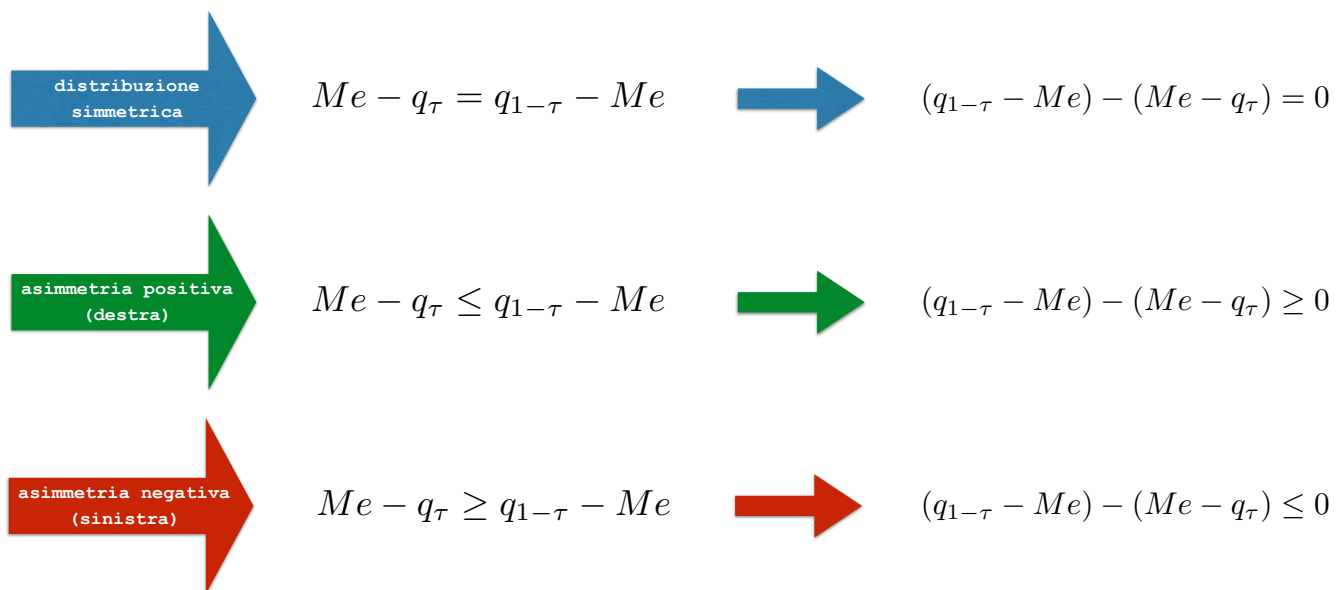
-512, -8, 0, 0, 0, +125, +125

Altre "forme" per
la forma: il ruolo
dei quantili

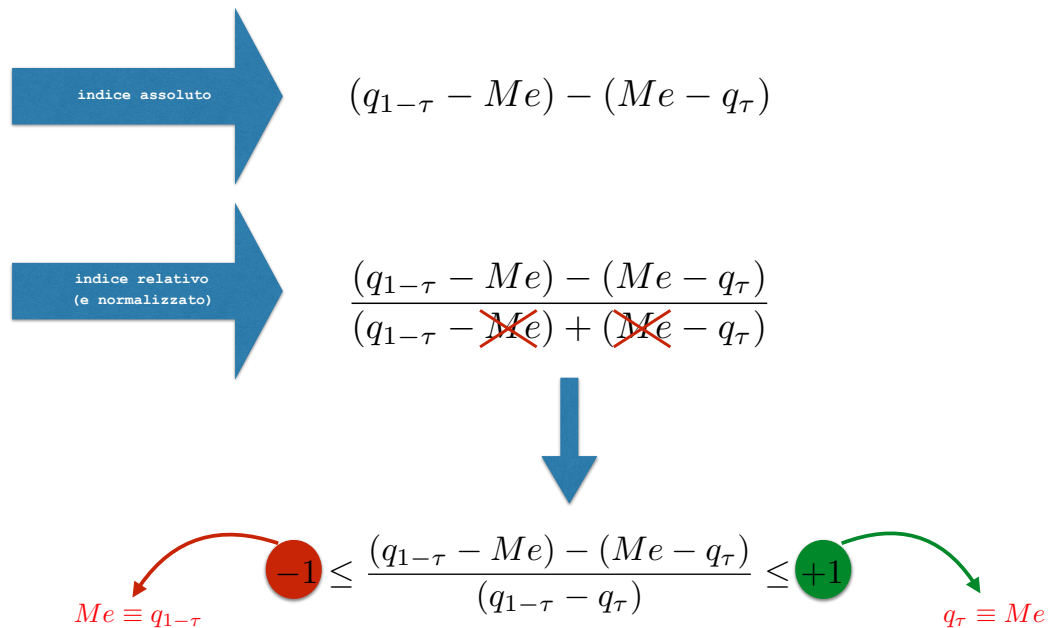
Il ruolo dei quantili



Il ruolo dei quantili



Il ruolo dei quantili: l'indice di Yule-Bowley



Il ruolo dei quantili: l'indice di Yule-Bowley

$$\tau = 0.25 \left\{ \begin{array}{l} q_\tau = q_{0.25} = q_1 \\ q_{1-\tau} = q_{0.75} = q_3 \end{array} \right.$$

indice di Yule-Bowley

$$\frac{(q_3 - Me) - (Me - q_1)}{(q_3 - Me) + (Me - q_1)} = \frac{(q_3 - Me) - (Me - q_1)}{(q_3 - q_1)}$$

Facciamo un po' di ordine?



Tre diverse approcci alla **forma**

scarti

Fisher

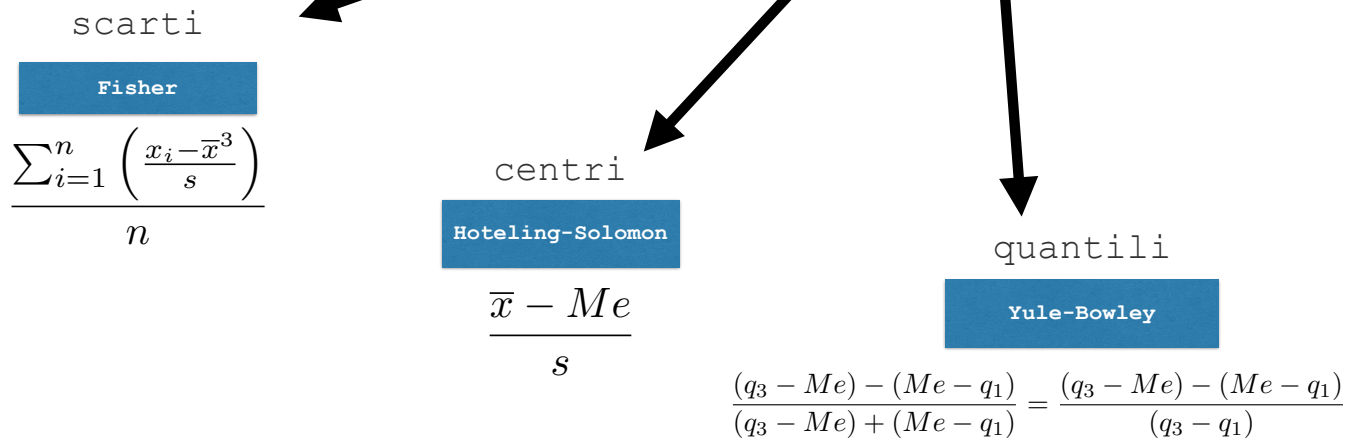
centri

Hotelling-Solomon

quantili

Yule-Bowley

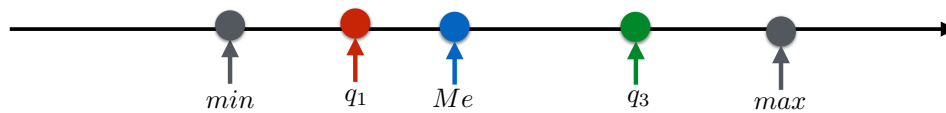
Tre diverse approcci alla **forma**



5 numeri "magici"
?

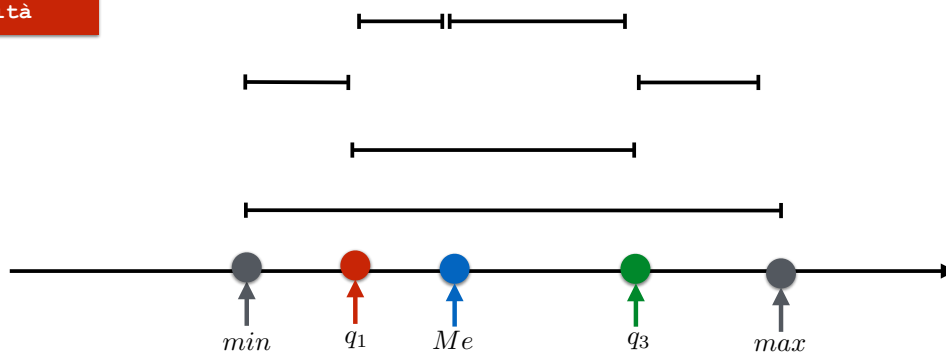
La sintesi a 5: min, q_1, Me, q_3, max

5 indici di posizione



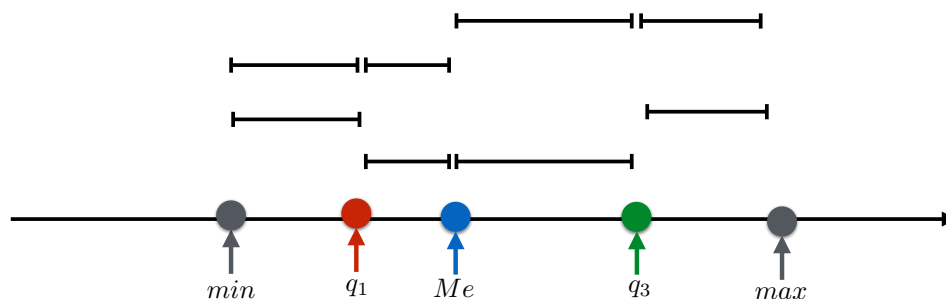
La sintesi a 5: min, q_1, Me, q_3, max

2 (e non solo) indici di variabilità

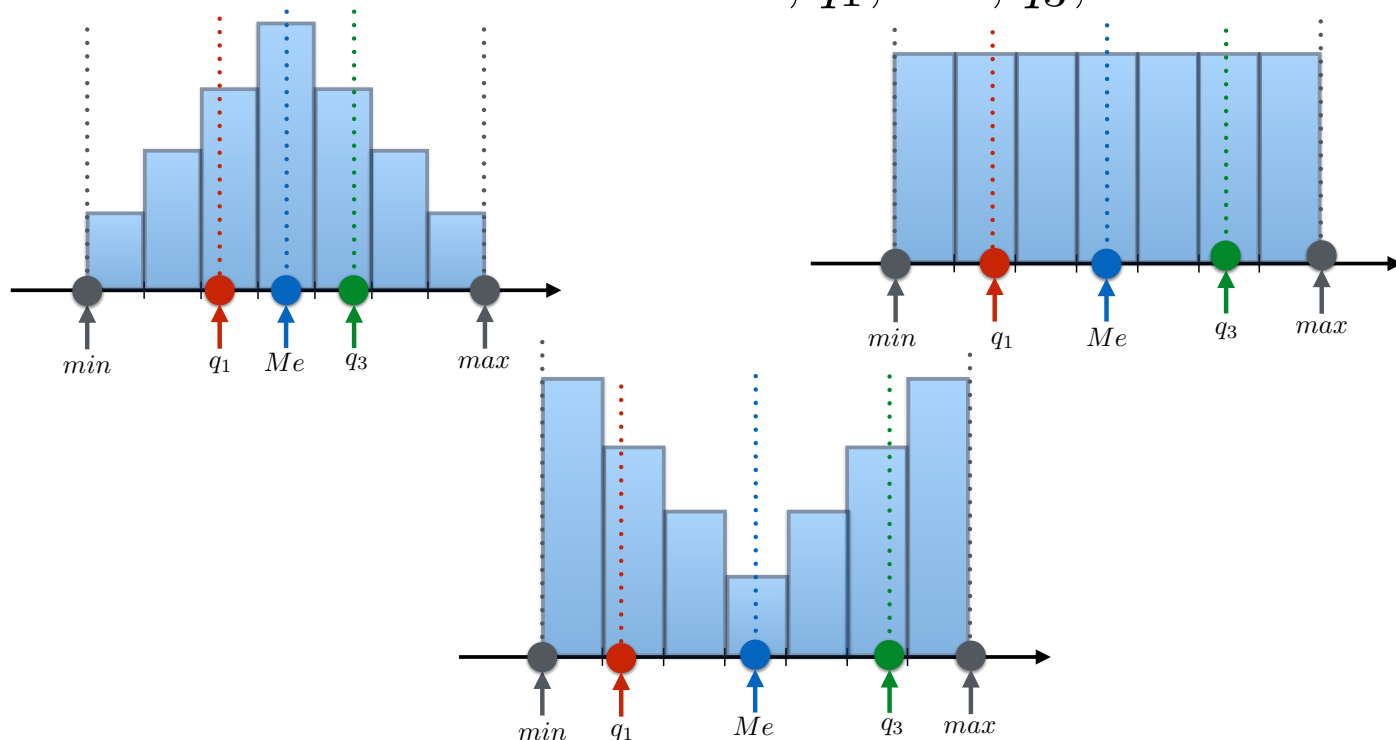


La sintesi a 5: $\min, q_1, Me, q_3, \max$

1 (e non solo) indice
di forma



La sintesi a 5: $\min, q_1, Me, q_3, \max$

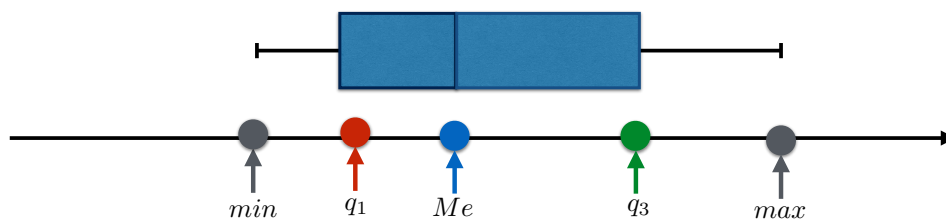


Una scatola coi baffi



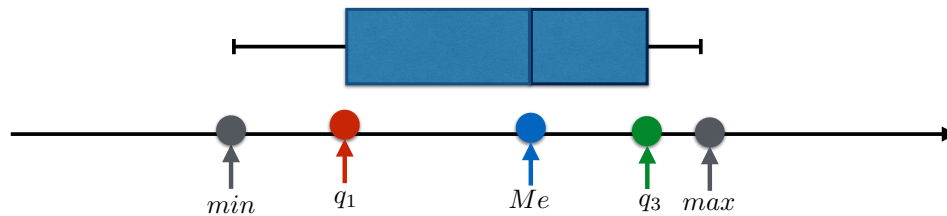
La sintesi a 5: il boxplot

Asimmetria positiva
(destra)



La sintesi a 5: il boxplot

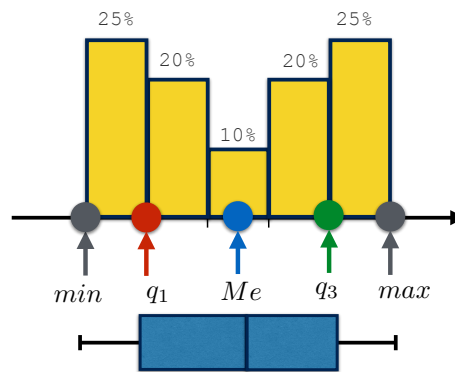
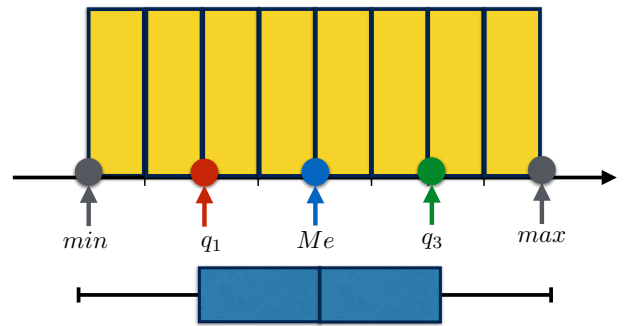
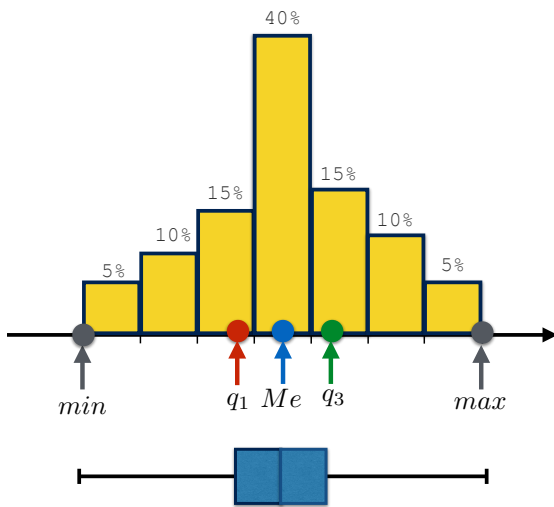
Asimmetria negativa
(sinistra)



Scatole per tutti
i gusti



La sintesi a 5: il boxplot



La sintesi a 5: il boxplot

