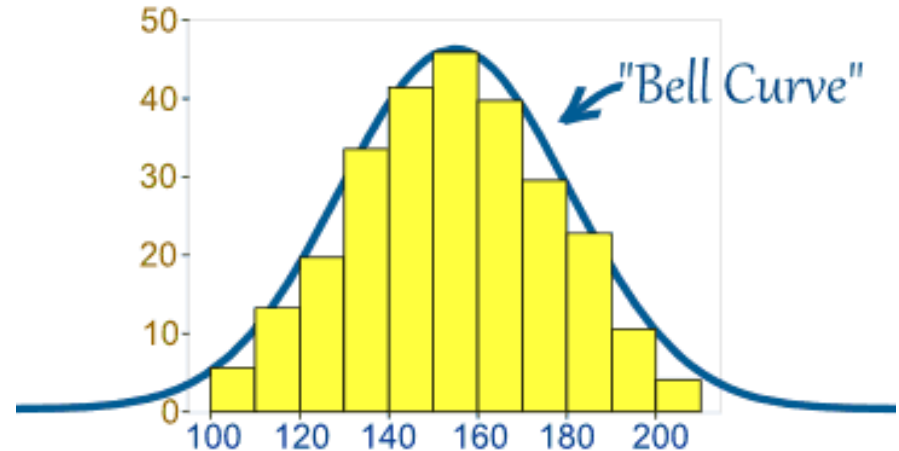
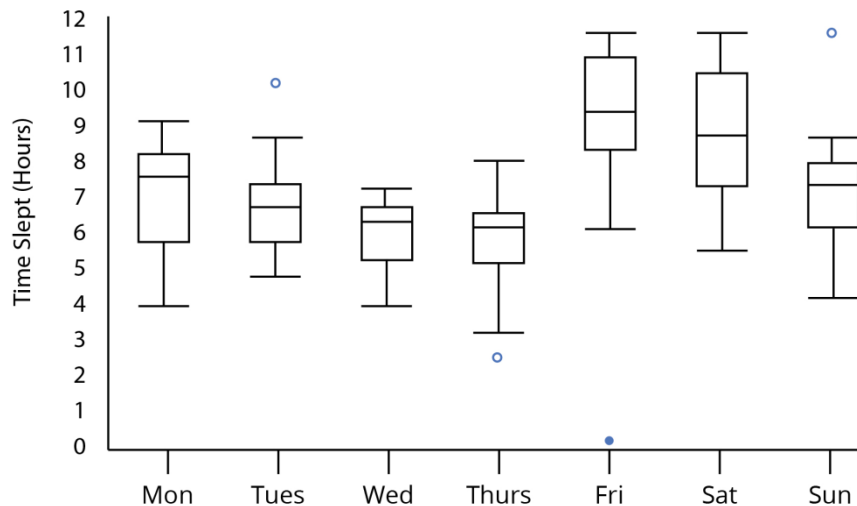


# Cenni di statistica

La **statistica** è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno in condizioni di incertezza o non determinismo, ovvero di non completa conoscenza di esso o parte di esso.

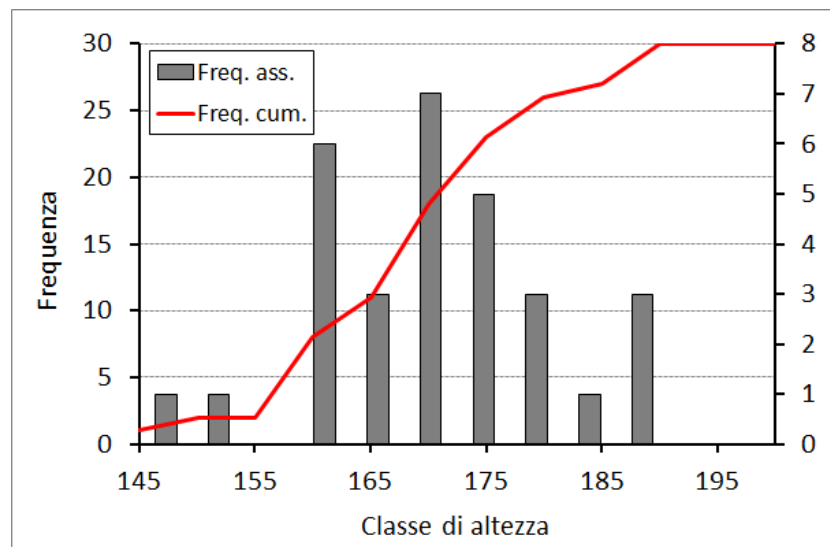
Essa è uno strumento del metodo scientifico e si avvale della matematica e del metodo sperimentale per studiare i modi in cui un fenomeno collettivo può essere sintetizzato e compreso. Ciò avviene attraverso la raccolta e l'analisi delle informazioni relative al fenomeno studiato



# Frequenze assolute e relative

ESEMPIO: altezza rilevata su 30 individui

classe	freq. ass.	freq. rel.	cumulate
145	1	3.33%	1
150	1	3.33%	2
155	0	0.00%	2
160	6	20.00%	8
165	3	10.00%	11
170	7	23.33%	18
175	5	16.67%	23
180	3	10.00%	26
185	1	3.33%	27
190	3	10.00%	30
195	0	0.00%	30
200	0	0.00%	30



**Frequenza assoluta:** numero di unità statistiche (individui) che presentano una data modalità (altezza).

**Frequenza relativa:** frazione o proporzione di unità statistiche che presentano tale modalità

individuo N°	altezza
1	150
2	170
3	172
4	172
5	157
6	180
7	172
8	172
9	173
10	170
11	167
12	165
13	167
14	180
15	180
16	185
17	190
18	158
19	145
20	165
21	190
22	190
23	160
24	160
25	167
26	168
27	156
28	166
29	158
30	162

Per identificare il numero minimo di classi (C) si utilizza la formula di Sturges:

$$C = 1 + \frac{10}{3} \log_{10} N$$

dove N è il numero di osservazioni. Nell'esempio precedente abbiamo N=30 quindi il numero di classi è 5.92 (approssimiamo a C=6)

L'ampiezza (a) di ciascuna classe sarà data quindi da  $a = \text{range}/C$ ;

esempio in classe: si fa un sondaggio sui presenti chiedendo come sono arrivati in Facoltà (a piedi, bicicletta, auto, vesuviana)

**Frequenza assoluta:** Su Excel selezionare le celle adiacenti alle classi di frequenza, usare la funzione FREQUENZA, selezionare matrice\_dati e matrice\_classi e premere contemporaneamente Ctrl+Shift+Enter

Appunti Carattere Allinear

F2 {=FREQUENZA(C2:C31,E2:E13)}

B	C	D	E	F
individuo N°	altezza		classe	freq. ass.
1	149		145	1
2	170		150	0
3	172		155	0
4	172		160	6
5	157		165	3
6	180		170	7
7	172		175	5
8	172		180	3
9	173		185	1
10	170		190	3
11	167		195	0
12	165		200	0
13	167			
14	180			
15	180			
16	185			
17	190			
18	158			
19	145			
20	165			
21	190			
22	190			
23	160			
24	160			
25	167			
26	168			
27	156			
28	166			
29	158			
30	162			

Argomenti funzione

FREQUENZA

Matrice\_dati C2:C31 = {149\170\172\172\157\180\172\172\173\...

Matrice\_classi E2:E13 = {145\150\155\160\165\170\175\180\185\...

= {1\1\0\6\3\7\5\3\1\3\0\0\0}

Calcola la frequenza con cui si presentano valori compresi in un intervallo e restituisce una matrice verticale di numeri con elemento in più rispetto a Matrice\_classi.

Matrice\_dati è una matrice o un riferimento ad un insieme di valori di cui si calcola la frequenza. Gli spazi e il testo vengono ignorati.

Risultato formula = 1

[Guida relativa a questa funzione](#)

OK Annulla



**Frequenza assoluta:** in alternativa, un metodo più comodo è andare su Dati, Analisi dati, Istogramma.

The screenshot shows the Microsoft Excel interface. The 'Dati' ribbon is highlighted with a red circle. The 'Analisi dati' dialog box is open, and 'Istogramma' is selected in the list of tools, also highlighted with a red circle. The background shows a data table with columns for absolute frequency, relative frequency, and cumulative frequency.

asse	freq. ass.	freq. rel.	cumulate
145	1	3.33%	1
150	1	3.33%	2
155	0	0.00%	2
160	6	20.00%	8
165	3	10.00%	11
170	7	23.33%	18
175	5	16.67%	23
180	3	10.00%	26
185	1	3.33%	27
190	3	10.00%	30
195	0	0.00%	30

Se Analisi dati non è attivo, bisogna attivarlo andando su File, Opzioni, Componenti aggiuntivi, Vai, spuntare Strumenti di analisi

The image shows the 'Opzioni di Excel' (Excel Options) dialog box, specifically the 'Componenti aggiuntivi' (Add-ins) section. The 'Componenti aggiuntivi' tab is selected in the left-hand menu. The main area displays a list of add-ins under the heading 'Visualizzazione e gestione dei componenti aggiuntivi'. The 'Componenti aggiuntivi di applicazioni attivi' (Active application add-ins) section is expanded, showing a list of installed add-ins. The 'Componente aggiuntivo Risolutore' (Solver Add-in) is highlighted. Below this list, the 'Componente aggiuntivo Risolutore' (Solver Add-in) is selected, and its details are shown in a separate window titled 'Componenti aggiuntivi' (Add-ins).

**Componenti aggiuntivi**

Componenti aggiuntivi disponibili:

- Componente aggiuntivo Risolutore
- Strumenti di analisi
- Strumenti di analisi - VBA
- Strumenti di conversione euro

**Componente aggiuntivo Risolutore**

Strumento per l'ottimizzazione delle formule e la risoluzione delle equazioni

OK Annulla Sfoglia... Automazione...

Opzioni di Excel

Generale  
Formule  
Strumenti di correzione  
Salvataggio  
Lingua  
Impostazioni avanzate  
Personalizzazione barra multifunzione  
Barra di accesso rapido  
**Componenti aggiuntivi**  
Centro protezione

Visualizzazione e gestione dei componenti aggiuntivi

**Componenti aggiuntivi**

Nome	Per
<b>Componenti aggiuntivi di applicazioni attivi</b>	
Componente aggiuntivo Risolutore	C:\...
PhantomPDF Creator COM Addin	C:\...
Send to Bluetooth	C:\F
Strumenti di analisi	C:\...
<b>Componenti aggiuntivi di applicazioni inattivi</b>	
Contenuto invisibile	C:\F
Data (XML)	C:\...
Dati XML personalizzati	C:\F
Fogli di lavoro nascosti	C:\F
Intestazioni e piè di pagina	C:\F
Microsoft Actions Pane 3	
Righe e colonne nascoste	C:\F
Strumenti di analisi - VBA	C:\...
Strumenti di conversione euro	C:\...

**Componenti aggiuntivi correlati a documenti**  
*Nessun componente aggiuntivo correlato a documenti*

**Componenti aggiuntivi di applicazioni disattivati**  
*Nessun componente aggiuntivo di applicazione disattivato*

Componente aggiuntivo: Componente aggiuntivo Risolutore  
Autore:  
Compatibilità: Informazioni sulla compatibilità non disponibili  
Percorso: C:\Program Files (x86)\Microsoft Office\Office14\Library\SOLVER\SOLVER.XLAM

Descrizione: Strumento per l'ottimizzazione delle formule e la risoluzione delle equazioni

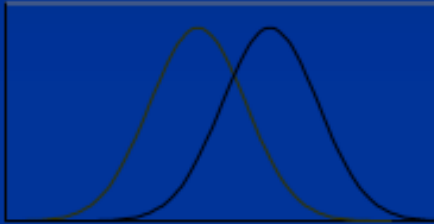
Gestisci: Componenti aggiuntivi di Excel

OK Annulla

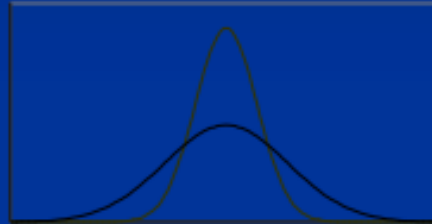
# Statistica descrittiva

## Indici caratteristici delle distribuzioni

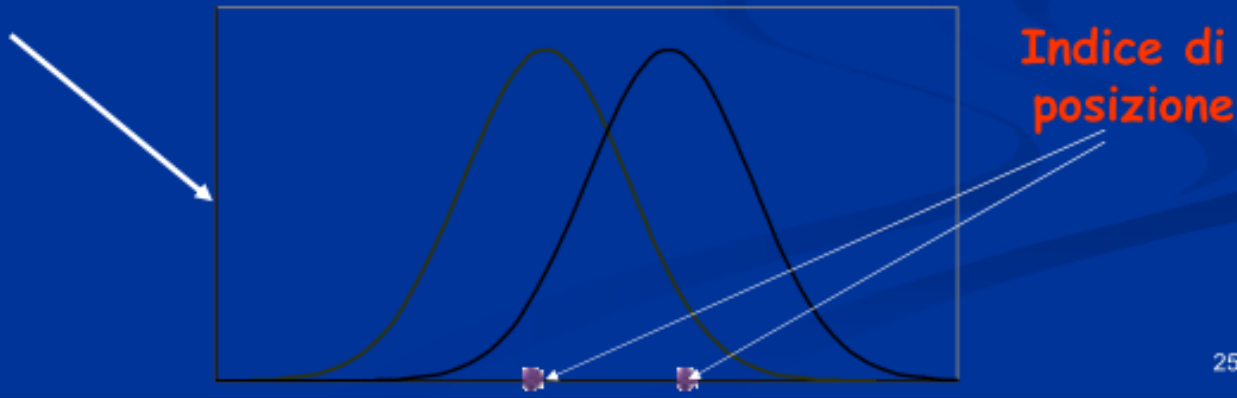
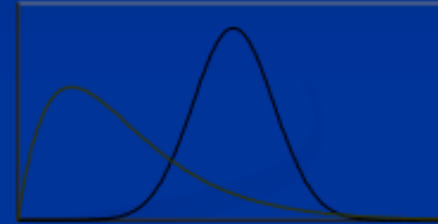
posizione



variabilità



forma



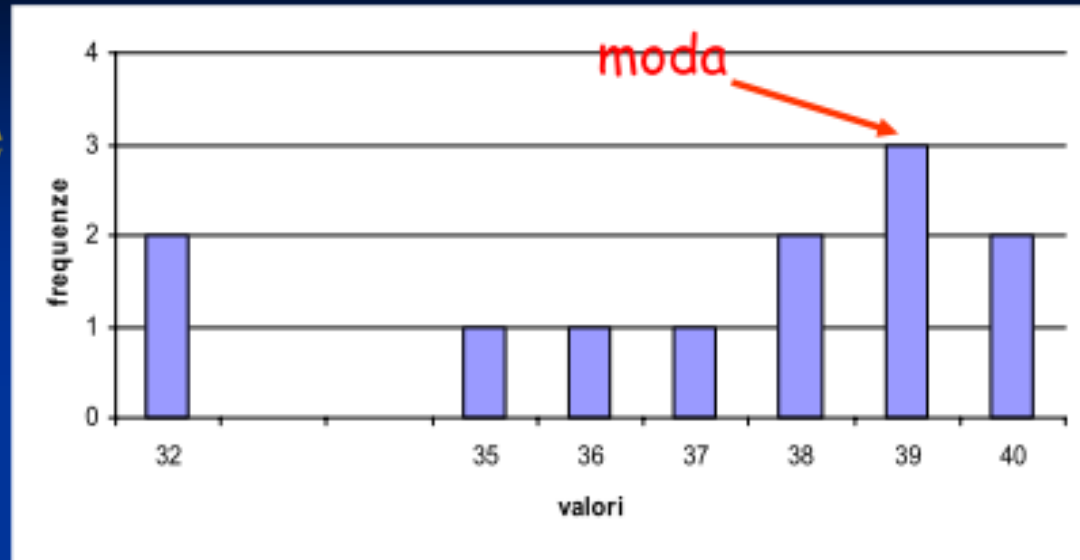
25

Indici di posizione: **moda**, **mediana**, **media**. Indicano la tendenza centrale di un insieme di dati (baricentro della distribuzione)

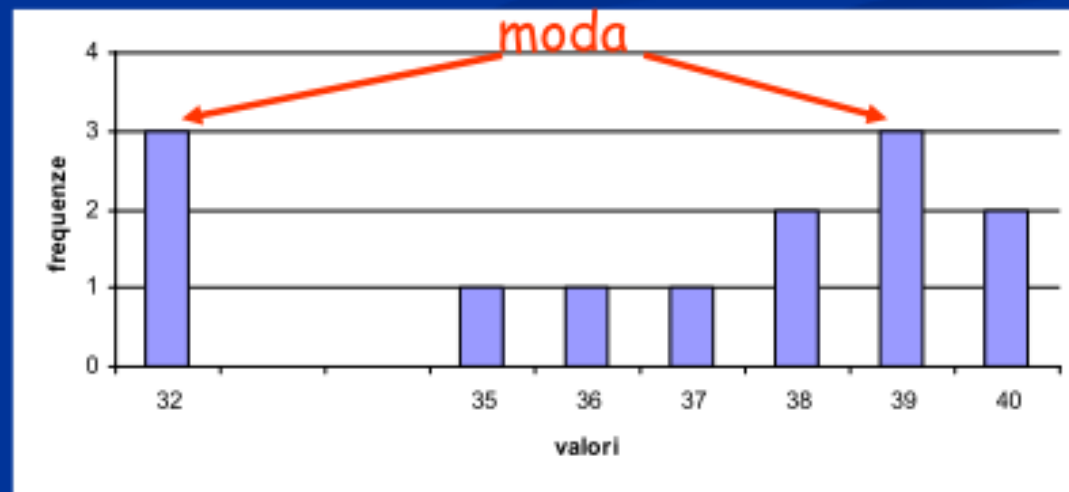
Moda: modalità a cui corrisponde la frequenza più alta e non è influenzata da valori estremi

## Moda

➤ unimodale



➤ bimodale



**Mediana:** valore che divide la distribuzione in due parti con numero uguale di termini al di sopra (50%) e al di sotto (50%) del suo valore. Non è influenzata da valori estremi.

Quantili

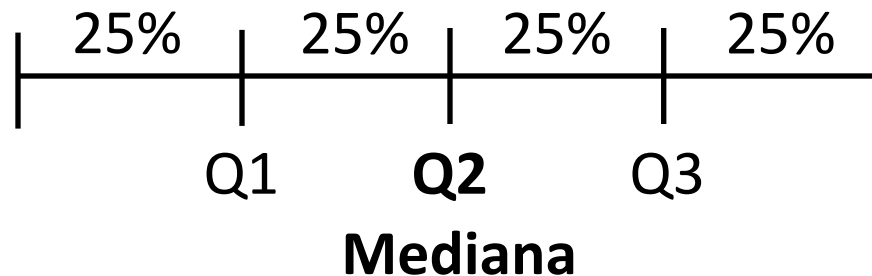
$p=0.5$  (mediana)

$p=0.25, 0.5, 0.75$  (quartili)

$p=0.1, 0.2, \dots, 0.8, 0.9$  (decili)

$p=0.01, 0.02, \dots, 0.98, 0.99$  (percentili)

Esempio dei quartili (Q1, Q2, Q3) che dividono la distribuzione in quattro porzioni di ugual numerosità



**Media:** indice di posizione più utilizzato ma è influenzato dai valori estremi

Media aritmetica

$$M_a = \frac{1}{n} \sum_{i=1}^n x_i$$

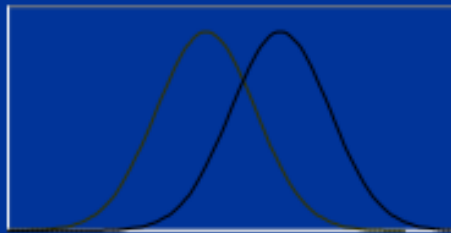
Media ponderata

$$M_{a,pond} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$$

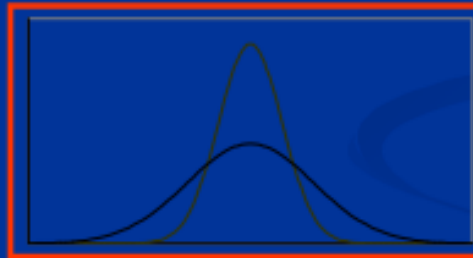
Proprietà della media: la sommatoria degli scarti dalla media è nulla (dimostra che la media è il baricentro della distribuzione di frequenza)

# Indici caratteristici delle distribuzioni

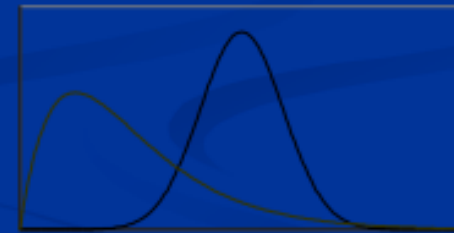
posizione



variabilità



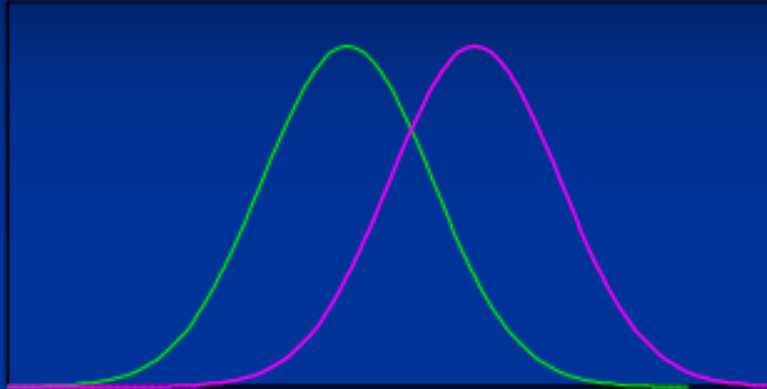
forma



54

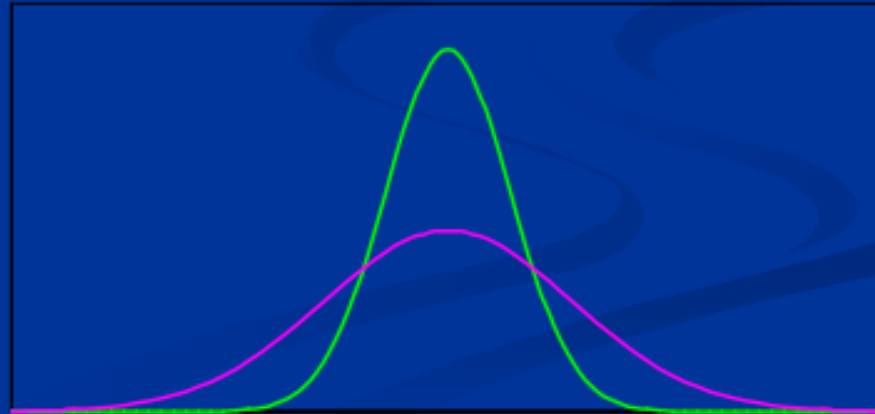
Gli indici di posizione non bastano a caratterizzare una distribuzione di dati, occorre informazione anche sulla dispersione intorno alla tendenza centrale

# Variabilità



media diversa,  
stessa variabilità

stessa media,  
variabilità diversa



55

L'indice di variabilità non assume valori negativi, il valore nullo è associato a variabilità nulla

# Indici di variabilità (o dispersione) (misure di variazione)

## ➤ indici assoluti:

- campo di variazione (range)
- scarto semplice medio
- varianza
- deviazione standard (scarto quadratico medio)

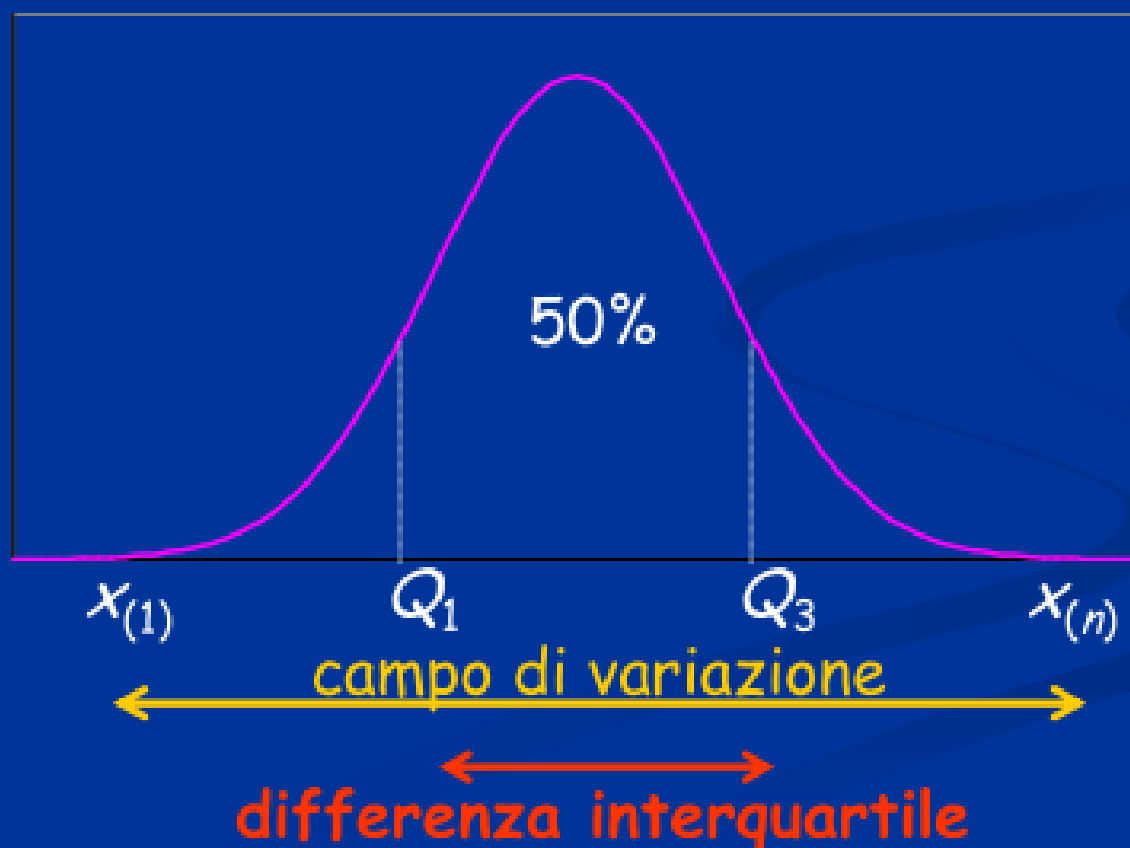
## ➤ indici relativi:

- coefficiente di variazione

# Campo di variazione (range)

Valore più alto  $-$  Valore più basso

$$R = x_{(1)} - x_{(n)}$$



# Varianza

## Devianza

(somma degli scarti al quadrato)

$$D = \sum (x_i - \bar{x})^2$$

## Varianza

(media degli scarti al quadrato)

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

# Deviazione standard (scarto quadratico medio)

**Deviazione standard**  
(scarto quadratico medio)

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

# Calcolo della deviazione standard

<i>alimento</i>	<i>energia kcal</i>	<i>xi-M</i>	<i>(xi-M)^2</i>
pane	276	-90.25	8145.06
grissini	433	66.75	4455.56
crackers	428	61.75	3813.06
fette	410	43.75	1914.06
biscotti	418	51.75	2678.06
pasta	356	-10.25	105.06
riso	362	-4.25	18.06
pizza	247	-119.25	14220.56
<b>Totale</b>	<b>2930</b>	<b>0.00</b>	<b>35349.50</b>

$$\bar{x} = 366.25$$

$$\sigma^2 = 4418.69$$

$$\sigma = \sqrt{4418.69} = 66.47$$

Per la partecipazione a una gara di matematica una scuola deve formare una squadra di 6 studenti; con una selezione preliminare, attraverso un test con un punteggio massimo di 100 punti, sulla base della media dei migliori punteggi risultano tre squadre a pari merito. Con quale criterio può essere scelta la squadra da mandare alla gara? Calcolare i valori di media, varianza e scarto quadratico medio con le funzioni di Excel (MEDIA, VAR, DEV.STD)

<b>squadra</b>	<b>punteggi degli studenti</b>					
<b>A</b>	73	76	77	85	88	90
<b>B</b>	74	74	78	84	88	91
<b>C</b>	72	77	79	82	84	95

<b>media squadra A</b>	81,5
<b>media squadra B</b>	81,5
<b>media squadra C</b>	81,5

<b>varianza squadra A</b>	49,9
<b>varianza squadra B</b>	52,7
<b>varianza squadra C</b>	61,1

<b>scarto quadratico medio squadra A</b>	7,06
<b>scarto quadratico medio squadra B</b>	7,26
<b>scarto quadratico medio squadra C</b>	7,82

# Coefficiente di variazione

$$V = \frac{\sigma}{\bar{x}}$$

$$V\% = \frac{\sigma}{\bar{x}} \times 100$$

Esempio precedente dove abbiamo  $\bar{x}=366.25$  e  $\sigma=66.47$

Pertanto otteniamo  $V\%=66.47/366.25 * 100 = 18.15\%$

# Esempio

Valori di VES (velocità di elettrosedimentazione, mm/ora) misurati su due gruppi di 7 pazienti

$$\{A\}: \{ 8, 5, 7, 6, 35, 5, 4 \}$$

$$\{B\}: \{ 11, 8, 10, 9, 17, 8, 7 \}$$

- stessa media  $\bar{A} = 10 = \bar{B}$
- in  $\{A\}$  i valori sono più dispersi che in  $\{B\}$ 
  - in  $\{A\}$  i valori sono inclusi tra 4 e 35
  - in  $\{B\}$  i valori sono inclusi tra 7 e 17

# Esempio

{A}:

$$D = 8^2 + 5^2 + \dots 4^2 - (8 + 5 + \dots 4)^2 / 7 = 1440 - 700 = 740$$

$$\sigma^2 = 740 / 6 = 123.33 \quad \sigma = 11.1$$

$$V\% = 100(11.1/10) = 111\%$$

{B}:

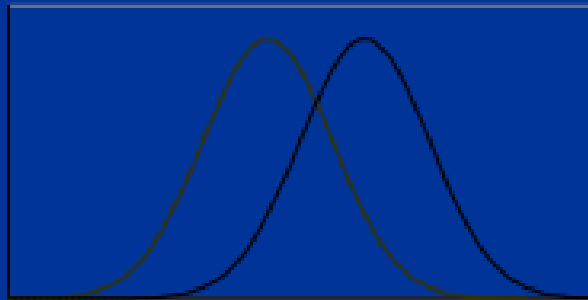
$$D = 11^2 + 8^2 + \dots 7^2 - (11 + 8 + \dots 7)^2 / 7 = 768 - 700 = 68$$

$$s^2 = 68 / 6 = 11.33 \quad \sigma = 3.4$$

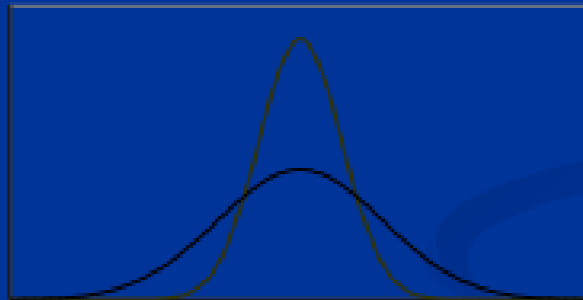
$$V\% = 100(3.4/10) = 34\%$$

# Indici caratteristici delle distribuzioni

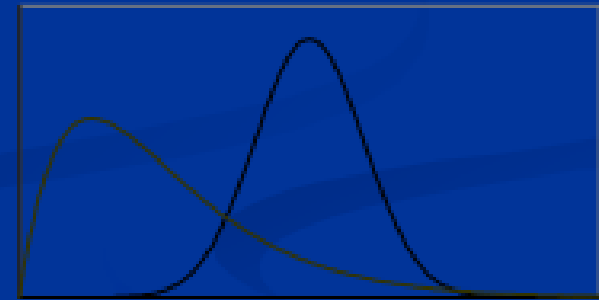
posizione



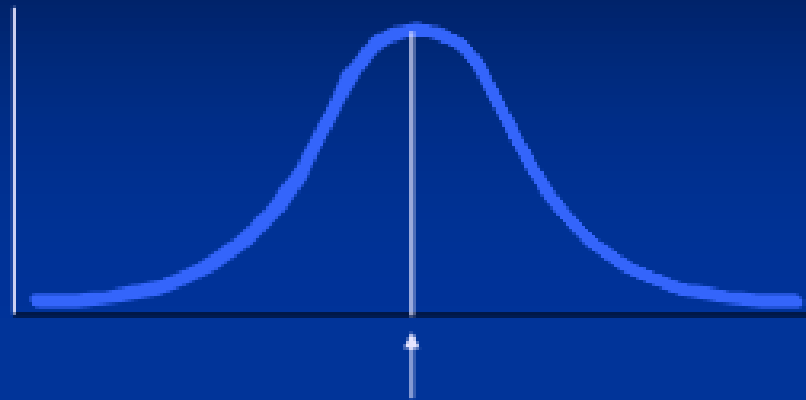
variabilità



forma

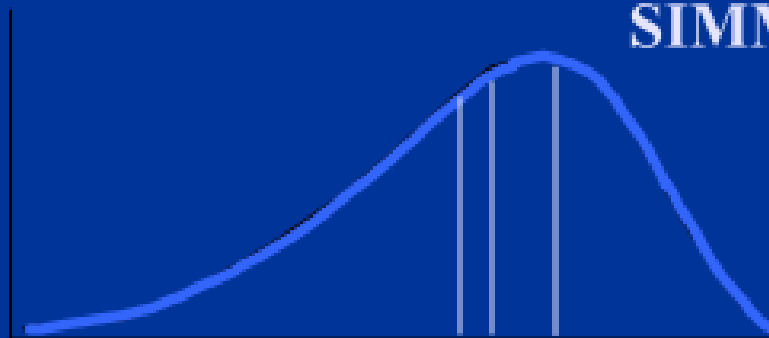


# Asimmetria (per distribuzioni unimodali)



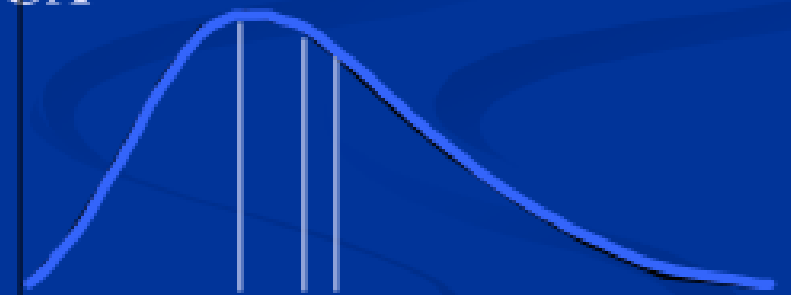
Moda = Media = Mediana

**SIMMETRICA**



Media ——— ↑ ↑ ↑ Moda  
                  Mediana

**ASIMMETRICA A SINISTRA**  
(negativa)



Moda ——— ↑ ↑ ↑ Media  
                  Mediana

**ASIMMETRICA A DESTRA**  
(positiva)

# Indici di asimmetria

$$\text{asimmetria} = \frac{\bar{x} - \text{moda}}{\sigma} \quad \text{asimmetria} = \frac{3(\bar{x} - \text{mediana})}{\sigma}$$

indici di asimmetria di Pearson:

$$\gamma_1 = \frac{\sum (x_i - \bar{x})^3 n_i}{n\sigma^3}$$

$$\beta_1 = \gamma_1^2$$

momento attorno alla media di ordine 3

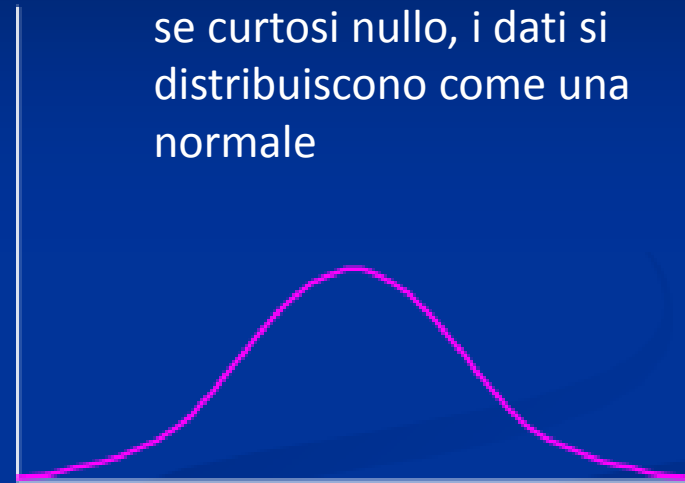
# Indici di curtosi

allungamento se  
curtosi maggiore  
di zero



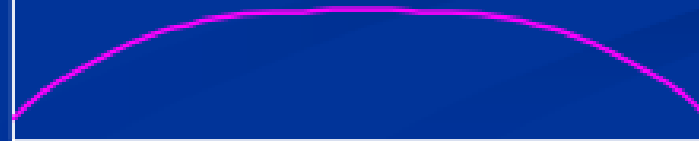
**leptocurtosi**

se curtosi nullo, i dati si  
distribuiscono come una  
normale



**mesocurtosi**

appiattimento se  
curtosi negativo



**platicurtosi**

$$\gamma_2 = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} - 3$$

# Indici: Schema riassuntivo

di posizione {

- **media:**  $\bar{x} = \frac{\sum_i x_i}{N}$
- **moda:** punto di max della distribuzione
- **mediana:** valore sotto al quale cadono la metà dei valori campionari. Si dispongono i dati in ordine crescente e si prende quello che occupa la posizione centrale ( $N$  dispari) o la media dei 2 valori in posizione centrale ( $N$  pari)

di dispersione {

- **varianza**  $\sigma^2 = \frac{\sum_i (x_i - \bar{x})^2}{N-1}$
- **deviazione standard**  $\sigma$
- **range**  $R = x_{\max} - x_{\min}$

>0 coda a ds  
 <0 coda a sin  
 =0 simmetrica

di di forma {

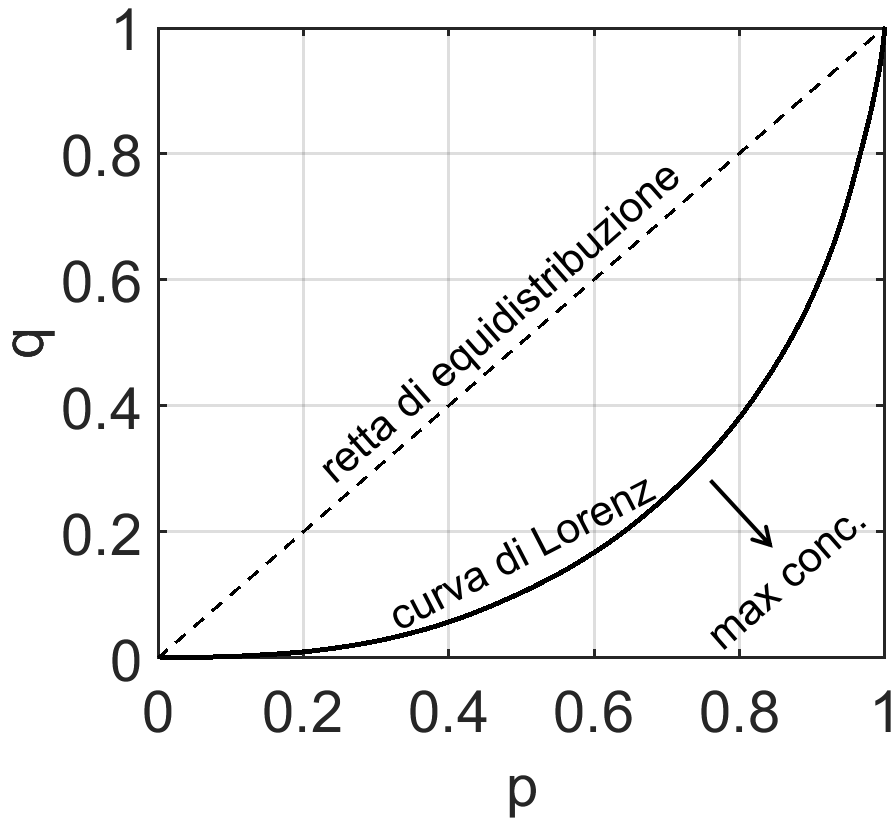
- **skewness (coeff. di asimmetria)**  $\frac{\sum_i \left( \frac{x_i - \bar{x}}{\sigma} \right)^3}{N}$

- **curtosi:** misura quanto la distribuzione è appuntita  
 >3 poco appuntita      <3 molto appuntita       $\frac{\sum_i \left( \frac{x_i - \bar{x}}{\sigma} \right)^4}{N}$



# La concentrazione

In presenza di  $i$ -unità (da 1 a  $n$ ), si consideri un carattere quantitativo e trasferibile da ordinare in senso crescente ( $x_i$ ) la cui somma è uguale a  $T$  (dove  $T=x_1+x_2+\dots+x_n$ ). La concentrazione esprime il modo in cui  $T$  si distribuisce fra le unità. Esempio di come il reddito cumulato si distribuisce tra i cittadini di una nazione.



Costruzione della curva di Lorenz:

- 1) ordinare in senso crescente i valori  $x_i$  ( $x_1 < x_2 < \dots < x_n$ )
- 2) calcolare le frequenze relative cumulate  $p$  ( $p_1=1/n$ ,  $p_2=2/n, \dots, p_n=n/n$ )
- 3) calcolare le proporzioni  $q$  ( $q_1=x_1/T$ ,  $q_2=(x_1+x_2)/T, \dots, q_n=T/T$ )

Il rapporto di concentrazione R si definisce come:

$$R = \frac{\sum_{i=1}^n (p_i - q_i)}{\sum_{i=1}^n p_i}$$

R è un indice normalizzato che varia tra:

- 1) R=0 (equidistribuzione)
- 2) R=1 (massima concentrazione)

