

# LEZIONI DI STATISTICA MEDICA

*Dott. SIMONE ACCORDINI*

## *Lezione n.8*

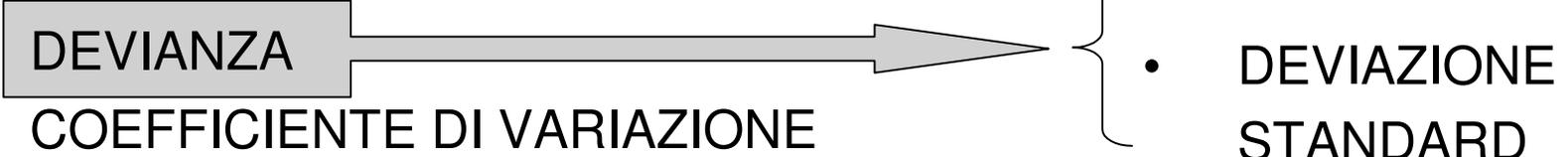
- Misure di dispersione*
- Misure di forma*



*Sezione di Epidemiologia & Statistica Medica  
Università degli Studi di Verona*

# MISURE DI DISPERSIONE

*(measures of dispersion)*

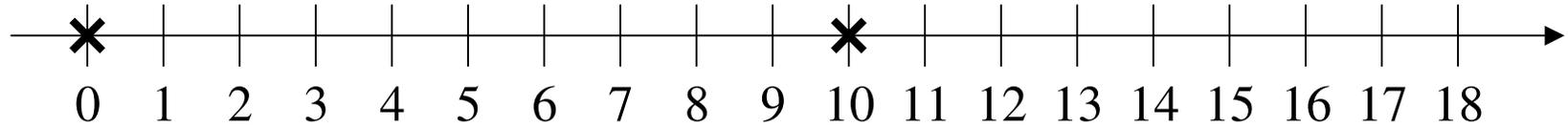
- CAMPO DI VARIAZIONE (range)
  - DISTANZA INTERQUARTILE
  - DEVIANZA
  - COEFFICIENTE DI VARIAZIONE
- 
- VARIANZA
  - DEVIAZIONE STANDARD

# RANGE (CAMPO DI VARIAZIONE)

$$\text{Range} = x_{\max} - x_{\min}$$

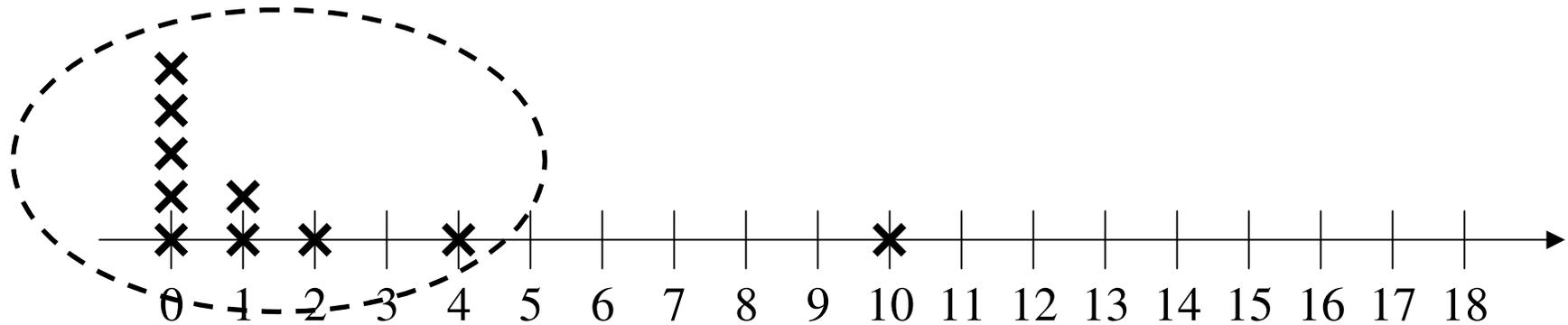
*differenza tra il valore massimo e il valore minimo osservato*

- Si basa soltanto sui valori estremi della distribuzione e non tiene conto dei valori intermedi.
- E' molto influenzato da osservazioni anomale (*outliers*).
- Tende ad aumentare al crescere del numero delle osservazioni.



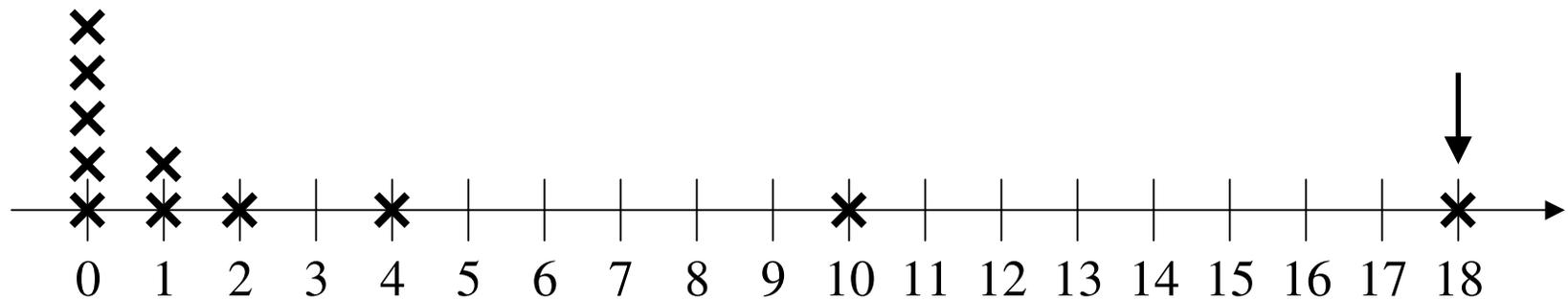
$n = 2 \rightarrow \text{Range} = x_{\max} - x_{\min} = 10 - 0 = 10$

num. linfonodi metastatici



$n = 10 \rightarrow \text{Range} = x_{\max} - x_{\min} = 10 - 0 = 10$

num. linfonodi metastatici



$n = 11 \rightarrow \text{Range} = x_{\max} - x_{\min} = 18 - 0 = 18$

num. linfonodi metastatici

# DISTANZA INTERQUARTILE

$$\text{IQR} = Q_3 - Q_1$$

*differenza tra il III° quartile ( $Q_3$ ) e il I° quartile ( $Q_1$ )*

- In questo intervallo ricade la metà dei valori osservati, posti esattamente al centro della distribuzione.
- Non è influenzata da osservazioni anomale o estreme.

## Statura matricole della Facoltà di Medicina (A.A. 95/96)

### MASCHI

$$\text{Range} = x_{\max} - x_{\min} = 193 - 162 = 31 \text{ cm}$$

Statura	Freq.	Cumul.
162	1	1
168	1	2
169	1	3
170	3	6
172	2	8
174	2	10
175	5	15
176	3	18
177	3	21
178	3	24
179	1	25
181	1	26
182	2	28
183	2	30
184	1	31
188	1	32
192	1	33
193	1	34
Totale	34	

Calcolo del I° quartile:

(rango percentile = 25)

$$\begin{aligned} 1. \text{ rango} &= (34+1) * 25 / 100 \\ &= 35 / 4 \approx 9 \end{aligned}$$

$$2. \text{ I}^\circ \text{ quartile: } Q_1 = 174 \text{ cm}$$

Calcolo del III° quartile:

(rango percentile = 75)

$$\begin{aligned} 1. \text{ rango} &= (34+1) * 75 / 100 \\ &= 35 * 3 / 4 \approx 26 \end{aligned}$$

$$2. \text{ III}^\circ \text{ quartile: } Q_3 = 181 \text{ cm}$$

$$\text{IQR} = Q_3 - Q_1 = 181 - 174 = 7 \text{ cm}$$

# DEVIANZA

Nella popolazione

dimensione della popolazione

$$\sum_{i=1}^N (x_i - \mu)^2$$

media nella popolazione (parametro)

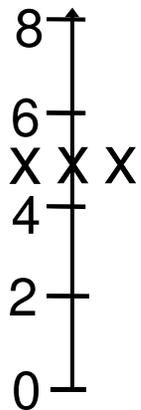
Nel campione

dimensione del campione

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

media nel campione (statistica)

- E' un indice di dispersione definito sulla base del concetto di SCARTO rispetto ad un punto centrale della distribuzione (media aritmetica).
- E' la base delle misure di dispersione per variabili quantitative (da essa discendono la VARIANZA e la DEVIATION STANDARD).



$$\bar{x} = 15 / 3 = 5$$

$$\sum x_i = 5 + 5 + 5 = 15$$

$$\text{devianza} = (5 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 = 0$$



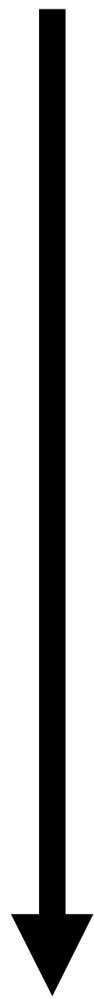
$$\sum x_i = 4 + 5 + 6 = 15$$

$$\text{devianza} = (4 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 = 2$$



$$\sum x_i = 2 + 5 + 8 = 15$$

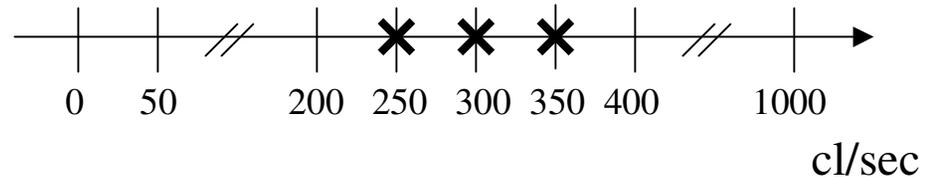
$$\text{devianza} = (2 - 5)^2 + (5 - 5)^2 + (8 - 5)^2 = 18$$



FEV <sub>1</sub>	n <sub>i</sub>	p <sub>i</sub>	(x <sub>i</sub> - $\bar{x}$ )	(x <sub>i</sub> - $\bar{x}$ ) <sup>2</sup>
250	1	0.33	-50	2500
300	1	0.33	0	0
350	1	0.33	50	2500
TOT	3	1	0	5000

$$\bar{x} = (250+300+350) / 3 = 300 \text{ cl/sec}$$

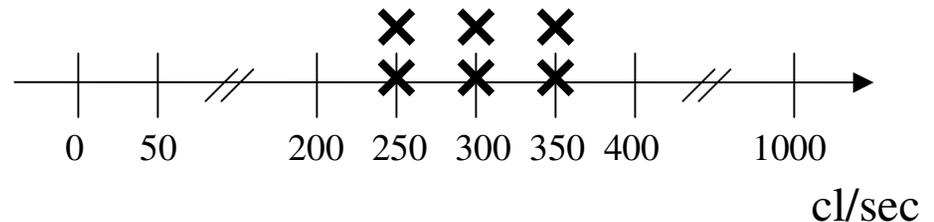
$$n = 3 \rightarrow \text{devianza} = 5000 \text{ cl}^2/\text{sec}^2$$



FEV <sub>1</sub>	n <sub>i</sub>	p <sub>i</sub>	(x <sub>i</sub> - $\bar{x}$ )n <sub>i</sub>	(x <sub>i</sub> - $\bar{x}$ ) <sup>2</sup> n <sub>i</sub>
250	2	0.33	-50*2	2500*2
300	2	0.33	0	0
350	2	0.33	50*2	2500*2
TOT	6	1	0	10000

$$\bar{x} = (250*2+300*2+350*2) / 6 = 300 \text{ cl/sec}$$

$$n = 6 \rightarrow \text{devianza} = 10000 \text{ cl}^2/\text{sec}^2$$



**La devianza raddoppia anche se la variabilità è costante!!!**

# VARIANZA

Nella popolazione

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

parametro

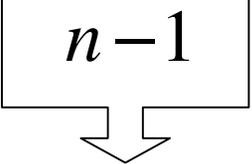
Nel campione (varianza corretta)

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

statistica

Gradi di  
Libertà

- E' una devianza media ossia la devianza rapportata al numero delle osservazioni campionarie ( $n$ ) o della popolazione ( $N$ ).
- E' la media aritmetica del quadrato degli scarti delle singole osservazioni dalla loro media aritmetica.

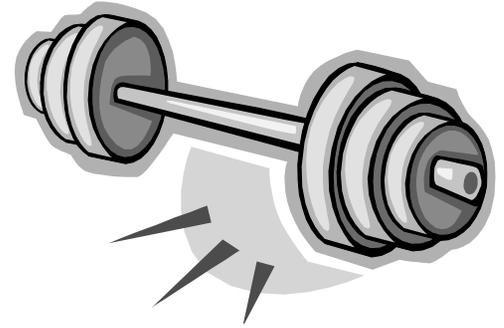
$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 / n}{n-1}$$


**VARIANZA:**  
formula per  
il calcolo

**I GRADI DI LIBERTÀ** rappresentano il numero di osservazioni indipendenti del campione, dal momento che sui dati disponibili è già stata calcolata una statistica (la media campionaria).

- Tiene conto di tutte le osservazioni ed è dunque influenzata da eventuali osservazioni anomale (*outliers*).
- Non è direttamente confrontabile con la media o altri indici di posizione in quanto l'unità di misura è elevata al quadrato.

# VARIANZA PONDERATA



Quando la distribuzione di frequenza è stata discretizzata in k classi

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

# Distribuzione di frequenza della statura delle matricole di Medicina dell'Università di Verona nell'A.A. 95/96

CLASSE	PUNTO CENTRALE ( $x_i$ )	FREQUENZA ASSOLUTA	$x_i * n_i$	$x_i^2 * n_i$
[150-155)	152.5	1	152.5 * 1 = 152.5	(152.5) <sup>2</sup> * 1 = 23256.25
[155-160)	157.5	8	157.5 * 8 = 1260.0	(157.5) <sup>2</sup> * 8 = 198450.00
[160-165)	162.5	24	162.5 * 24 = 3900.0	(162.5) <sup>2</sup> * 24 = 633750.00
[165-170)	167.5	34	5695.0	953912.50
[170-175)	172.5	27	4657.5	803418.75
[175-180)	177.5	19	3372.5	598618.75
[180-185)	182.5	9	1642.5	299756.25
[185-190)	187.5	1	187.5	35156.25
[190-195]	192.5	2	385.0	74112.50
TOTALE		125	21252.5	3620431.25

$$S^2 = \frac{\sum_{i=1}^k x_i^2 n_i - (\sum_{i=1}^k x_i n_i)^2 / n}{n - 1} = \frac{3620431.25 - (21252.5)^2 / 125}{124} = 57.1 \text{ cm}^2$$



# DEVIAZIONE STANDARD

Nella popolazione

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Nel campione (dev. st. corretta)

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

- Ha sempre un valore positivo.
- E' una misura della dispersione della variabile intorno alla media.
- E' direttamente confrontabile con le misure di posizione, essendo calcolata con la stessa unità di misura.

# Distribuzione di frequenza della statura delle matricole di Medicina dell'Università di Verona nell'A.A. 95/96

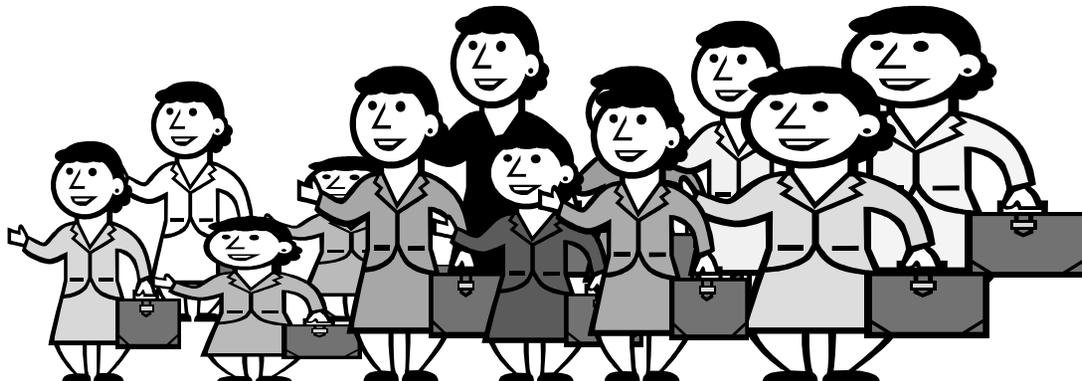
CLASSE	PUNTO CENTRALE ( $x_i$ )	FREQUENZA ASSOLUTA	$x_i * n_i$	$x_i^2 * n_i$
[150-155)	152.5	1	152.5 * 1 = 152.5	(152.5) <sup>2</sup> * 1 = 23256.25
[155-160)	157.5	8	157.5 * 8 = 1260.0	(157.5) <sup>2</sup> * 8 = 198450.00
[160-165)	162.5	24	162.5 * 24 = 3900.0	(162.5) <sup>2</sup> * 24 = 633750.00
[165-170)	167.5	34	5695.0	953912.50
[170-175)	172.5	27	4657.5	803418.75
[175-180)	177.5	19	3372.5	598618.75
[180-185)	182.5	9	1642.5	299756.25
[185-190)	187.5	1	187.5	35156.25
[190-195]	192.5	2	385.0	74112.50
TOTALE		125	21252.5	3620431.25

$$S = \sqrt{\frac{\sum_{i=1}^k x_i^2 n_i - (\sum_{i=1}^k x_i n_i)^2 / n}{n - 1}} = \sqrt{\frac{3620431.25 - (21252.5)^2 / 125}{124}} = \sqrt{57.1} = 7.6 \text{ (cm)}$$



# ESERCIZIO

*I dati seguenti si riferiscono al livello di emoglobina (X) in g/100 ml misurato in un campione di 70 donne:*



Determinate il range e la distanza interquartile della distribuzione (dati individuali).

9	11,4	12,9
9,3	11,4	13
9,4	11,4	13,1
9,7	11,5	13,1
10,2	11,6	13,2
10,2	11,6	13,3
10,3	11,7	13,3
10,4	11,7	13,4
10,4	11,8	13,4
10,5	11,8	13,5
10,6	11,9	13,5
10,6	11,9	13,6
10,7	12	13,7
10,8	12	13,7
10,8	12,1	14,1
10,9	12,1	14,6
10,9	12,1	14,6
10,9	12,2	14,7
11	12,3	14,9
11	12,5	15
11,1	12,5	
11,1	12,7	
11,2	12,9	
11,2	12,9	
11,3	12,9	

# ESERCIZIO

*I dati seguenti si riferiscono al livello di emoglobina ( $X$ ) in g/100 ml misurato in un campione di 70 donne:*

CLASSE	PUNTO CENTRALE ( $x_i$ )	FREQUENZA ASSOLUTA ( $n_i$ )
[9-10)	9.5	4
[10-11)	10.5	14
[11-12)	11.5	19
[12-13)	12.5	14
[13-14)	13.5	13
[14-15]	14.5	6
TOTALE		70



Determinate la varianza e la deviazione standard della distribuzione

*(dati raggruppati in intervalli di classe*

*-> media ponderata).*

**In alcune situazioni il confronto della variabilità all'interno di due gruppi di osservazioni è fuorviante se si utilizza la deviazione standard.**

### **Due gruppi con valori medi molto distanti**

Tre neonati pesano rispettivamente **3, 4 e 5 Kg** (media = **4 Kg**; dev.st. = **1 Kg**).

Tre bambini di 1 anno pesano **10, 11 e 12 Kg** (media = **11 Kg**; dev.st. = **1 Kg**).

**La deviazione standard è uguale nei due gruppi, ma il buon senso suggerisce che la variabilità del peso sia maggiore nei neonati.**

- La variabile misurata è la stessa ma i valori medi delle osservazioni nei due gruppi sono molto distanti (*le osservazioni nei due gruppi sono su diversi ordini di grandezza*)



## Due variabili diverse:

In 91 ragazze matricole di Medicina a Verona nell'A.A. 95/96,  
la media del **peso** era pari a **55.1 Kg** e la deviazione standard era pari a **5.7 Kg**,  
la media della **statura** era pari a **166.1 cm** e la deviazione standard era pari a **6.1 cm**.

**E' maggiore la variabilità del peso o la variabilità della statura?**

- Le variabili misurate nei due gruppi sono diverse (*le osservazioni nei due gruppi sono espresse con diverse unità di misura*)



# COEFFICIENTE DI VARIAZIONE (CV)

La deviazione standard viene espressa in percentuale della media.

$$CV = (\text{deviazione standard} / \text{media}) * 100$$

	Media	Dev. standard	CV
Neonati	4 Kg	1 Kg	<b>25.0 %</b>
Bambini 1 anno	11 Kg	1 Kg	<b>9.1 %</b>

**La variabilità del peso è maggiore nei neonati.**

	Media	Dev. standard	CV
Peso	55.1 Kg	5.7 Kg	<b>10.3 %</b>
Statura	166.1 cm	6.1 cm	<b>3.7 %</b>

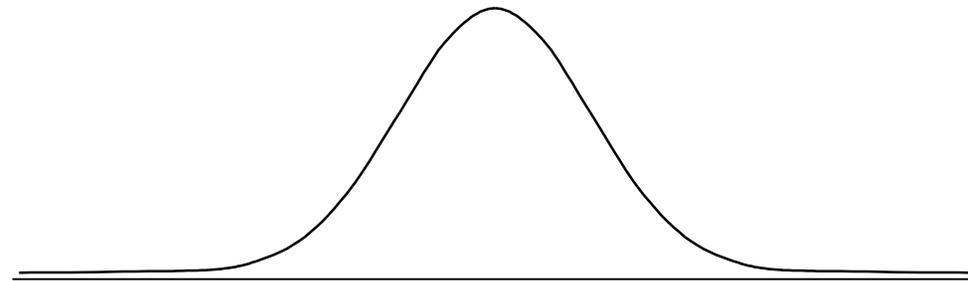
**La variabilità del peso è maggiore della variabilità della statura.**

# MISURE DI FORMA

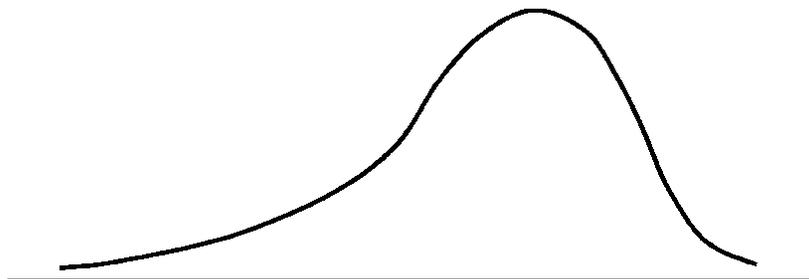
- **Misure di Simmetria (skewness)**

Indice di asimmetria:

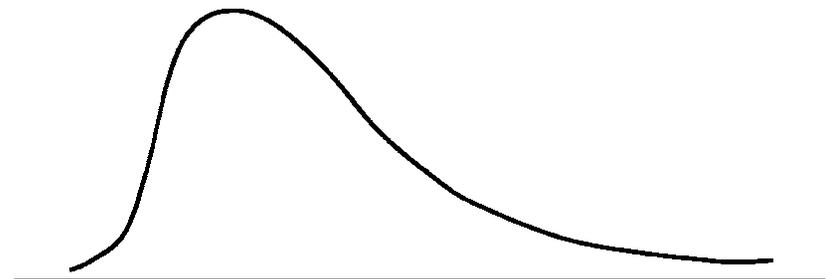
$$\frac{[\sum(x_i - \bar{x})^3]}{[\sum(x_i - \bar{x})^2]^{3/2}}$$



**Simmetria (indice = 0)**



**Asimmetria negativa (indice < 0)**

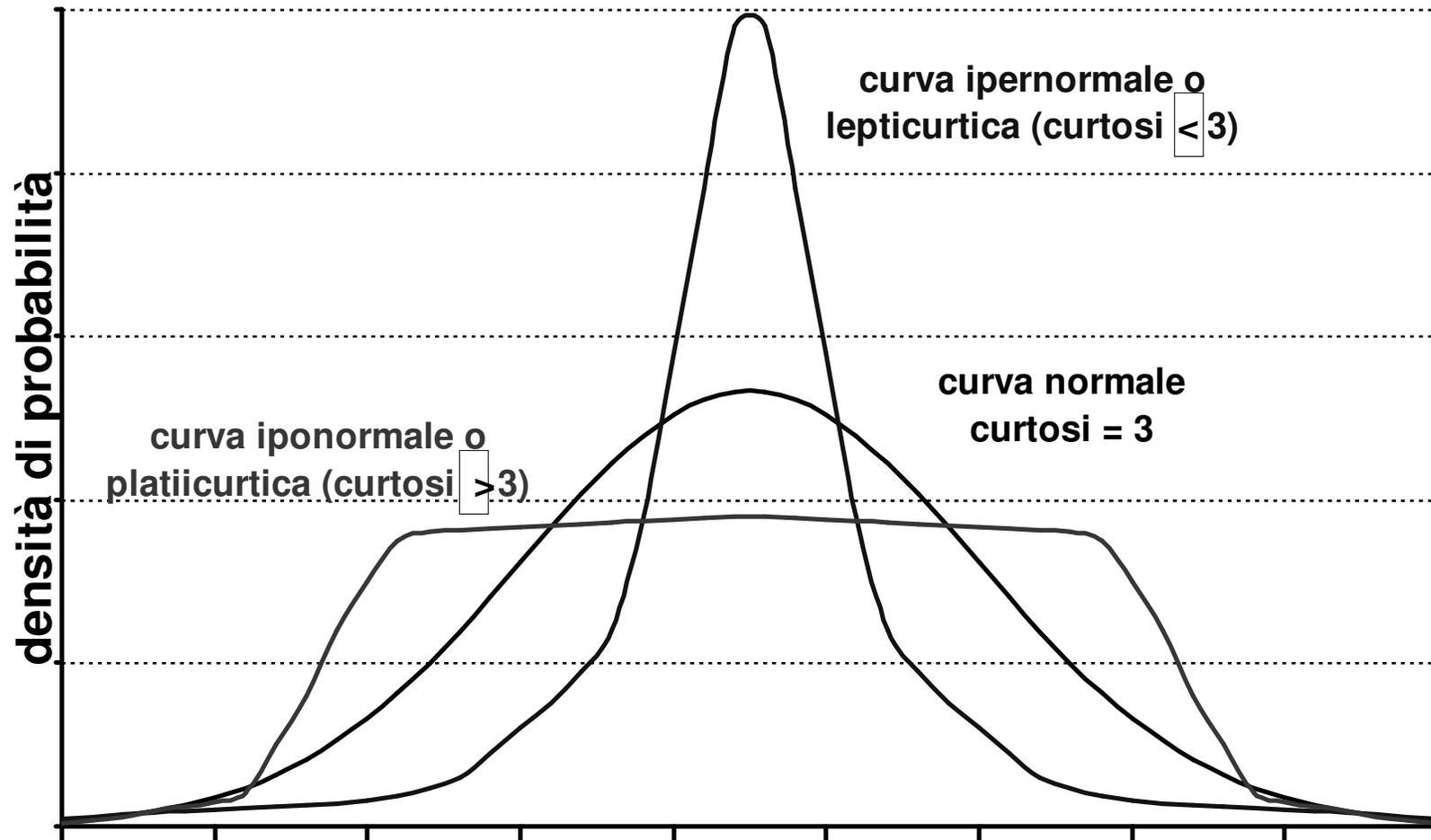


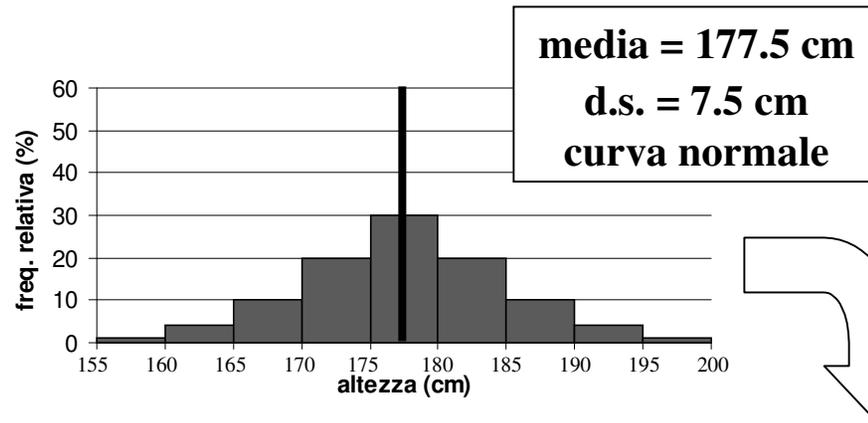
**Asimmetria positiva (indice > 0)**

# • Misure di Appiattimento o di Curtosi (kurtosis)

Indice di curtosi:

$$\frac{[\sum(x_i - \bar{x})^4]}{[\sum(x_i - \bar{x})^2]^2}$$





## POSIZIONE E SIMMETRIA

## DISPERSIONE

