




DATA LAB

GUARDA AVANTI

Big Data, nuove competenze
per nuove professioni.



"Anticipare la crescita con le nuove competenze sui Big Data" Operazione Rif. PA 2023-19167/RER approvata con DGR n° 843 del 29 maggio 2023 e co-finanziata dal Fondo Sociale Europeo Plus 2021-2027 Regione Emilia-Romagna



ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Operazione Rif. PA 2023-19167/RER/10/1, "ANTICIPARE LA CRESCITA CON LE NUOVE COMPETENZE SUI BIG DATA", approvata dalla Regione Emilia-Romagna con DGR n° 843 del 29/05/2023 e co-finanziata dal Fondo Sociale Europeo Plus 2021-2027

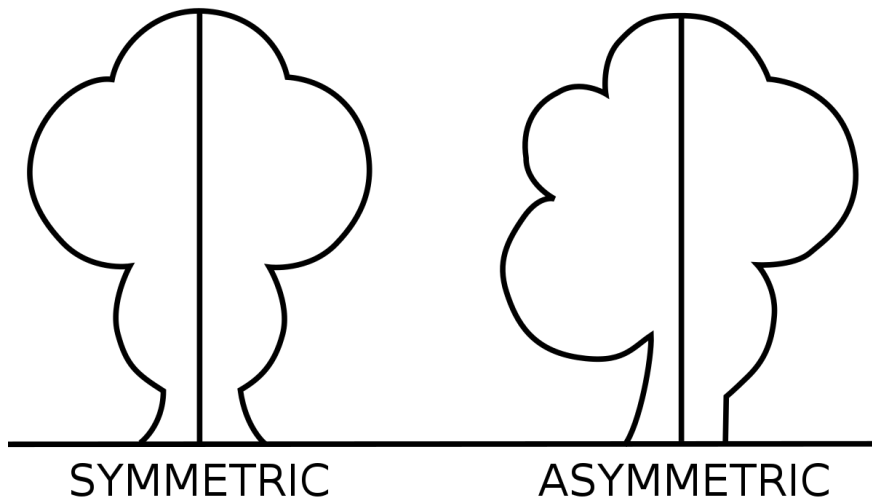
DATA LAB

Distribuzioni simmetriche e distribuzioni asimmetriche

- La forma delle distribuzioni di frequenza
- Misure di asimmetria: la asimmetria
- Misure di forma: la curtosi

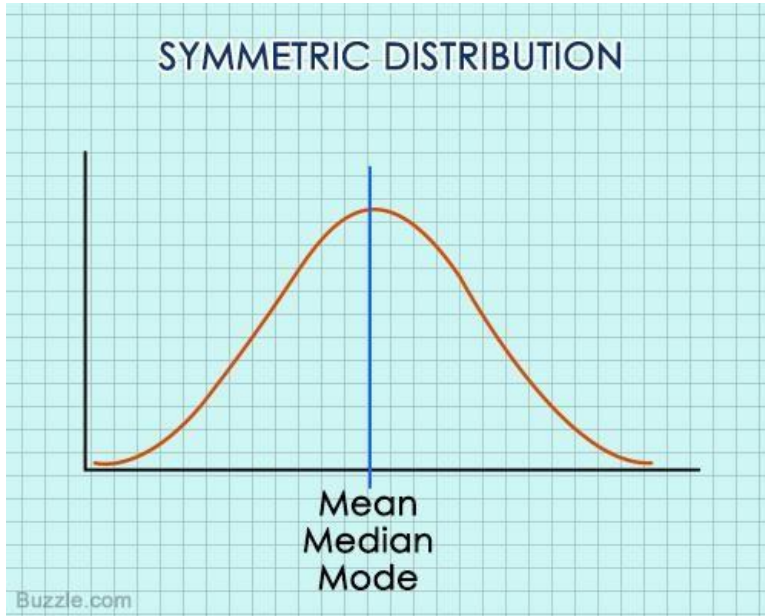
Distribuzioni simmetriche ed asimmetriche

- La simmetria



Distribuzioni simmetriche

- **Distribuzione simmetriche**

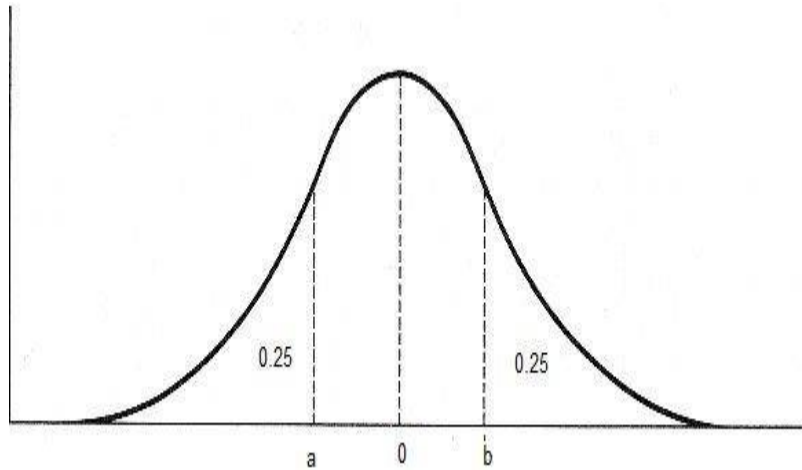


Nelle distribuzioni simmetriche:

Media, Moda e Mediana sono tutte perfettamente coincidenti (o molto vicine tra loro)

Forma normale di una distribuzione

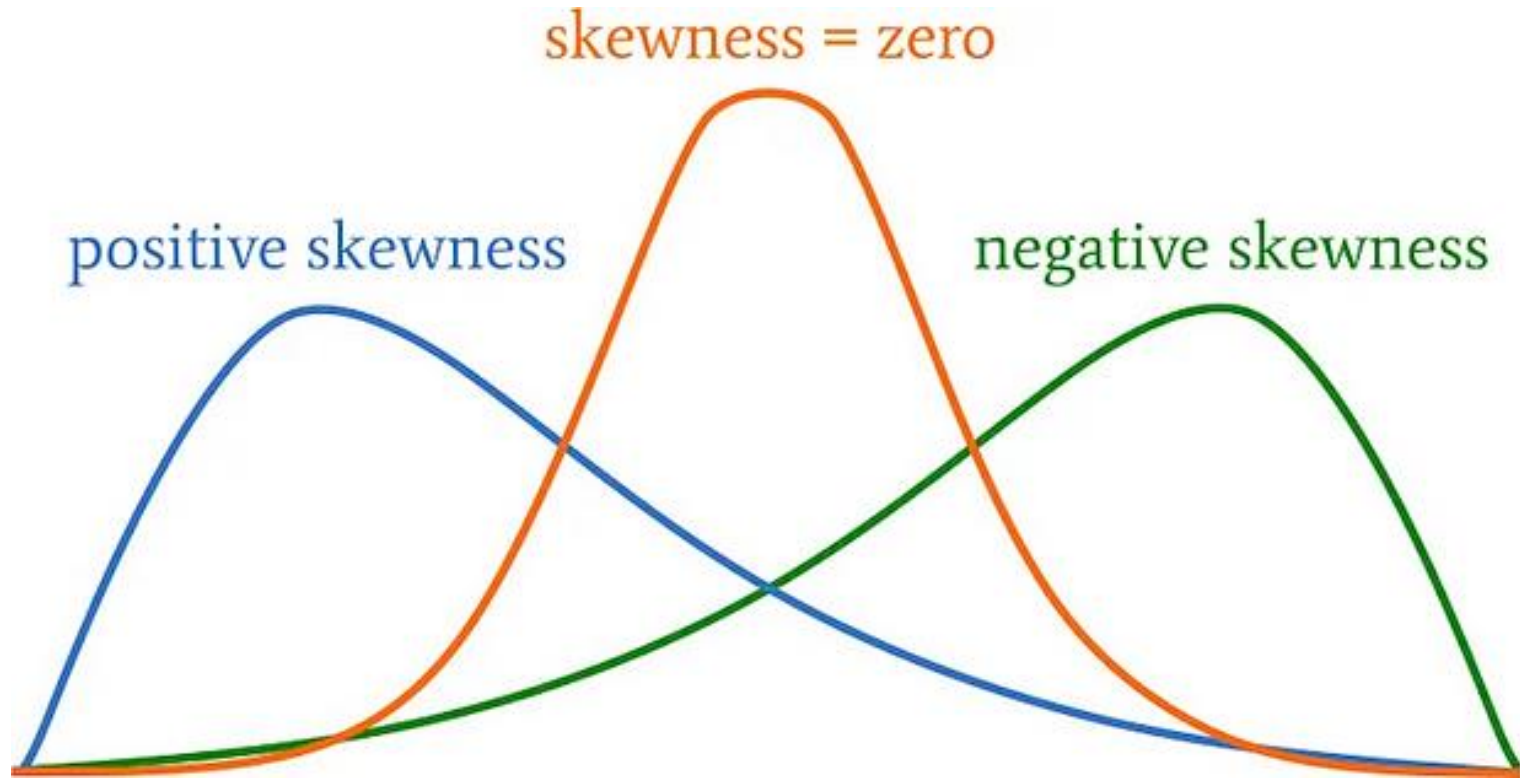
- **Distribuzione Normale**



- Presenta una forma campanulare

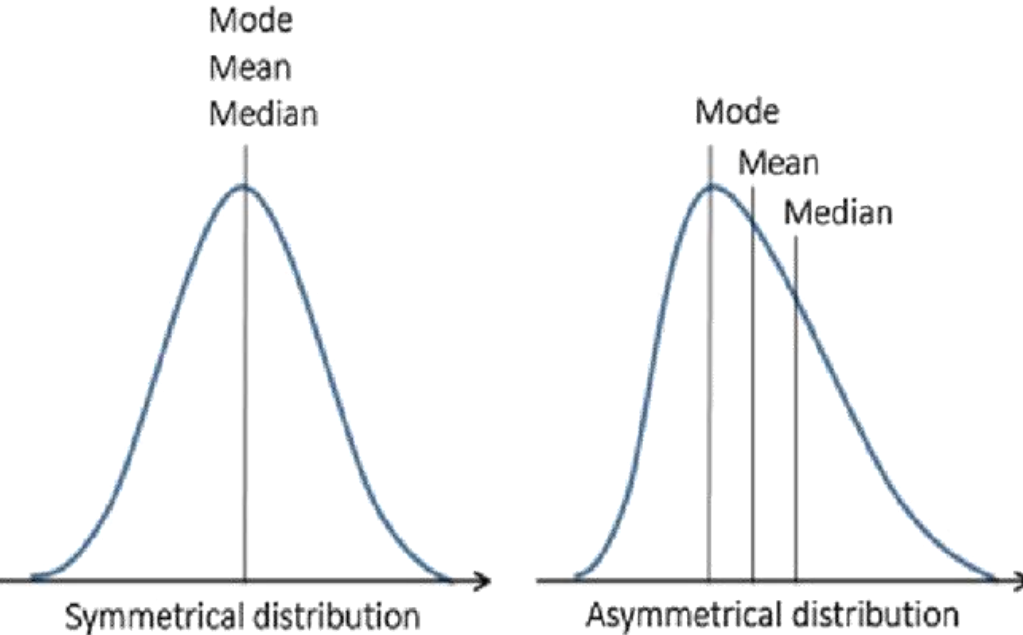
- E' simmetrica rispetto agli
indici di posizione
Media=Mediana=Moda

Skewness



Distribuzioni asimmetriche

- **Distribuzioni asimmetriche**

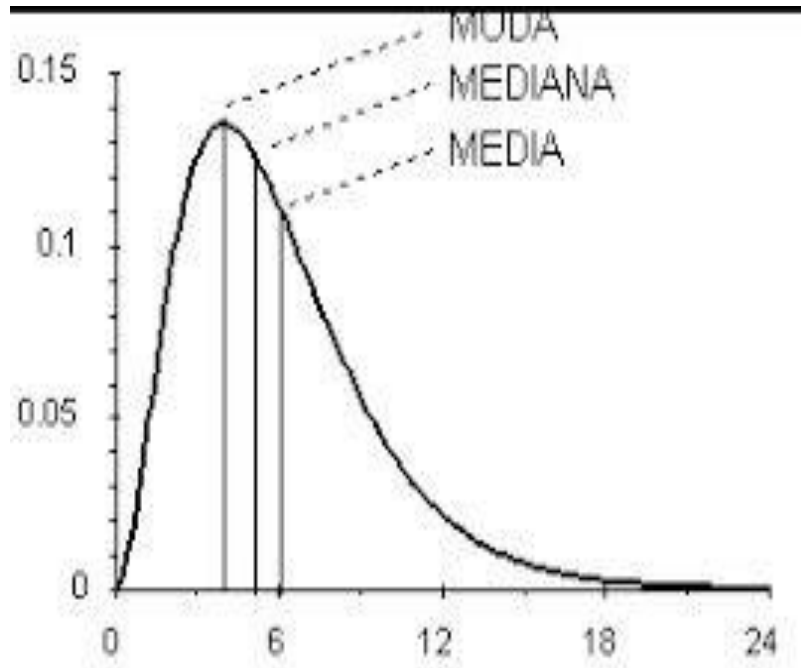


Nelle distribuzioni
asimmetriche:

Media, Moda e Mediana
non coincidono e la
distribuzione ha una
“gobba” rivolta a destra o a
sinistra

Di solito sono distribuzioni
non normali

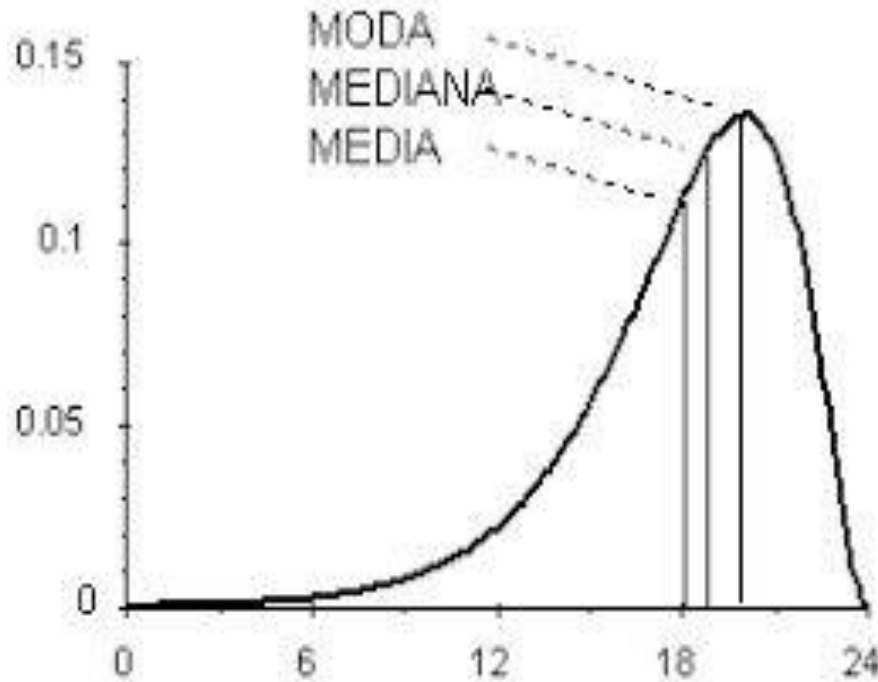
Asimmetria positiva



La distribuzione è caratterizzata da una coda allungata verso destra

Media > Mediana > Moda

Asimmetria negativa



La distribuzione è caratterizzata da una coda allungata verso sinistra

Media < Mediana < Moda

Come individuare la presenza di asimmetria

Un metodo empirico per individuare la presenza di asimmetria è confrontare gli indici di posizione della distribuzione considerata

Distribuzioni simmetriche

- Esempio

Classi	Frequenze assolute	Valori centrali (xi)	xi*ni
0-10	1	$10/2=5$	$5*1=5$
10-20	5	$30/2=15$	$5*15=75$
20-30	13	$50/2=25$	$13*25=325$
30-40	19	$70/2=35$	$19*35=665$
40-50	13	$90/2=45$	$13*45=585$
50-60	5	$110/2=55$	$5*55=275$
60-70	1	$130/2=65$	$65*1=65$
totale	57		1995

$$\mathbf{M=1995/57=35}$$

Distribuzioni simmetriche

- Esempio

Classi	Frequenze assolute	Frequenze relative cumulate
0-10	1	1/57=0.02
10-20	5	6/57=0.11
20-30	13	19/57=0.33
30-40	19	38/57=0.67
40-50	13	51/57=0.89
50-60	5	56/57=0.98
60-70	1	57/57=1
totale	57	

Classe - mediana: 30-40

$$\text{Mediana} = 30 + [(0.5 - 0.33) * (40 - 30) / (0.67 - 0.33)] = \mathbf{35}$$

Distribuzioni simmetriche

- Esempio

Classi	Frequenze assolute	Ampiezza di classe	Densità di frequenza
0-10	1	10	0.1
10-20	5	10	0.5
20-30	13	10	1.3
30-40	19	10	1.9
40-50	13	10	1.3
50-60	5	10	0.5
60-70	1	10	0.1
totale	57		

Classe-Modale: 30-40

Moda: $(30+40)/2=35$

Dato che Media=Mediana=Moda=35

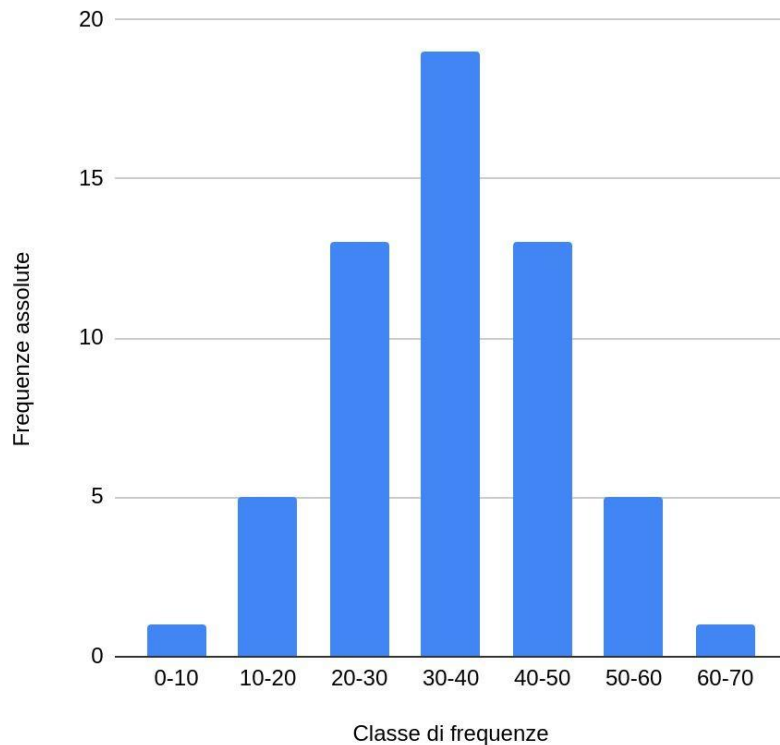
->La distribuzione è **SIMMETRICA**

Distribuzioni simmetriche

- Istogramma

Classe di frequenze	Frequenze assolute
0-10	1
10-20	5
20-30	13
30-40	19
40-50	13
50-60	5
60-70	1

Distribuzione simmetrica



Distribuzioni asimmetriche (II)

- Esempio

Classi	Frequenz e assolute ni	Valori centrali xi	xi*ni
0-10	1	5	5
10-20	10	15	150
20-30	21	25	525
30-40	16	35	560
40-50	8	45	360
50-60	5	55	275
60-70	1	65	65
Totale	62		1940

$$M = 1940/62 = \mathbf{31.29}$$

Distribuzioni asimmetriche (II)

- Esempio

Classi	Frequenze assolute	Ampiezza di classe	Densità di frequenz a
0-10	1	10	0.1
10-20	10	10	1
20-30	21	10	2.1
30-40	16	10	1.6
40-50	8	10	0.8
50-60	5	10	0.5
60-70	1	10	0.1

Classe-Modale: classe **20-30**

$$\mathbf{Mo} = (20 + 30) / 2 = \mathbf{25}$$

Distribuzioni asimmetriche (II)

- Esempio

Classi	Frequenze assolute	Frequenze cumulate relative
0-10	1	1/62=0.016
10-20	10	11/62=0.18
20-30	21	32/62=0.52
30-40	16	48/62=0.77
40-50	8	56/62=0.9
50-60	5	61/62=0.98
60-70	1	62/62=1
Totale	62	

Classe-Mediana: classe
20-30

$$Me = 20 + [(0.5 - 0.18) * (30 - 20) / (0.52 - 0.18)] = \mathbf{29.41}$$

Media > Mediana > Moda

La distribuzione è

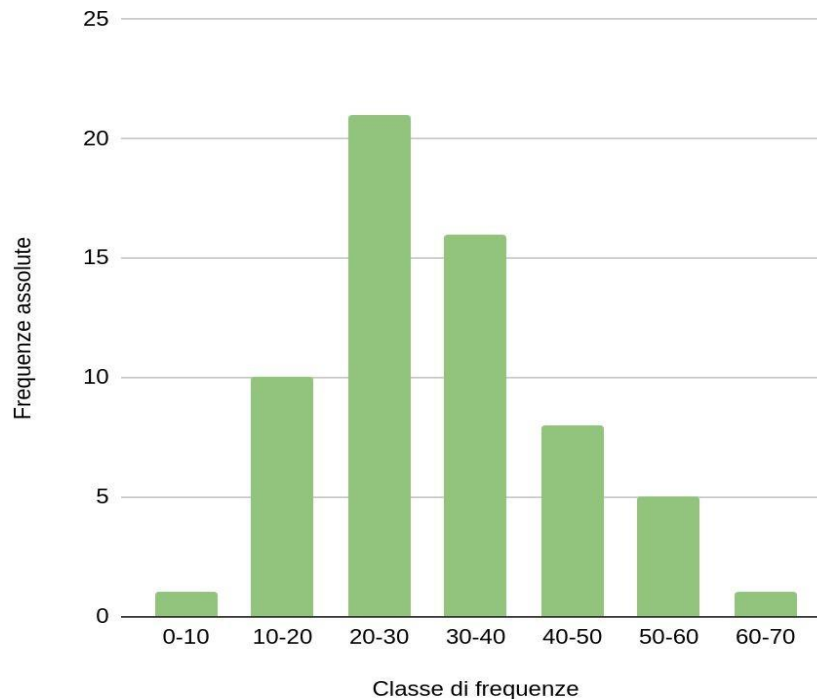
ASIMMETRICA

Distribuzioni asimmetriche

- Esempio

Classe di frequenze	Frequenze assolute
0-10	1
10-20	10
20-30	21
30-40	16
40-50	8
50-60	5
60-70	1

Distribuzione asimmetrica



Misure di asimmetria: i momenti

I momenti di ordine k una distribuzione sono definiti dalla seguente formula:

$$m_k = \sum_{i=1}^n \frac{1}{n} (x_i - M)^k$$

In questa formula M è la media e k è una potenza intera.

Se **$k=1$** L'equazione diventa la somma degli scarti attorno alla media M (vale 0 per costruzione)

Se **$k=2$** la formula calcola la varianza

Misure di asimmetria

Le misure di asimmetria o skewness

La asimmetria è un rapporto tra il

$$g_1 = \frac{m_3}{m_2^{3/2}}$$

Momento di ordine 3 .

m_3

E il *momento di ordine 2*
 m_2 elevato $3/2$

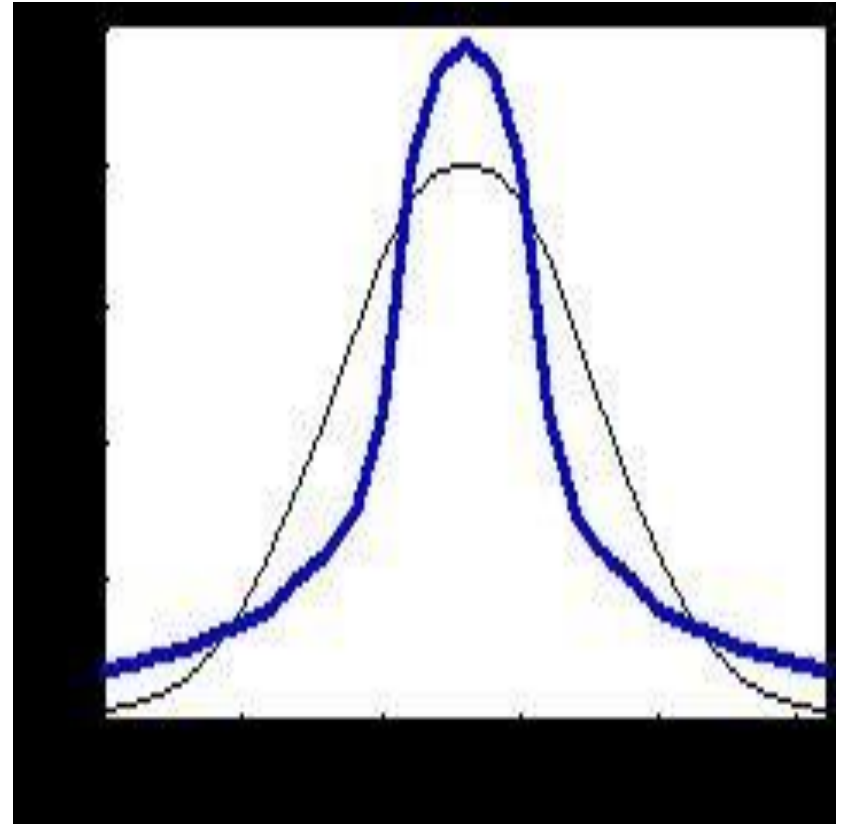
$$g_1 = \frac{\sum_{i=1}^n \frac{1}{n} (x_i - M)^3}{\sqrt[3]{\left(\sum_{i=1}^n \frac{1}{n} (x_i - M)^2 \right)^2}} = \frac{m_3}{m_2^{3/2}}$$

Curtosi

Se una distribuzione è simmetrica o quasi simmetrica allora la distribuzione può essere più o meno **appuntita** o più o meno **appiattita** rispetto alla distribuzione normale (o di Gauss)

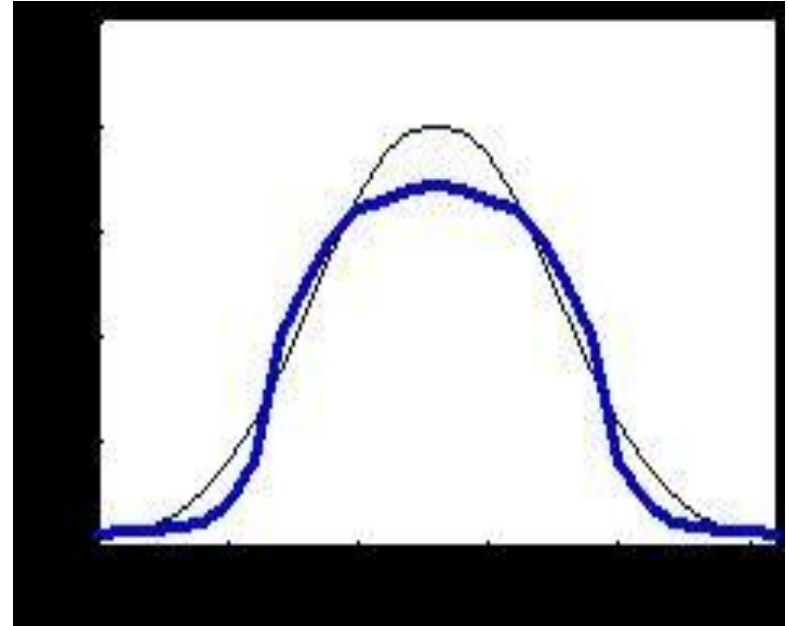
Curva Leptocurtica

Se la distribuzione è più
appuntita rispetto alla
distribuzione normale si dice
LEPTOCURTICA

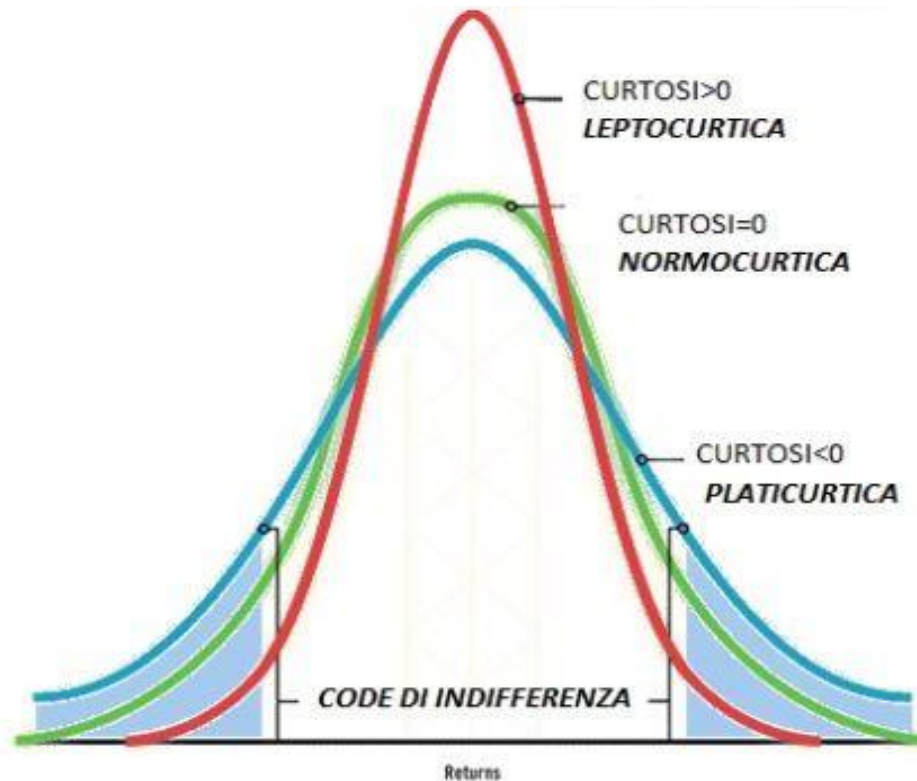


Curva Platicurtica

Se la distribuzione è più **piatta** rispetto alla distribuzione normale si dice **PLATICURTICA**



Confronto delle curtosi



Indice di Pearson-Curtosi (I)

Si calcola come media aritmetica delle quarte potenze della variabile standardizzata Z

Può assumere 3 valori:

-Se è pari a 3 la distribuzione assume una forma NORMALE

-Se è maggiore di 3 la distribuzione ha una forma appuntita rispetto alla normale-> la distribuzione è LEPTOCURTICA

-Se è inferiore a 3 la distribuzione ha una forma appiattita rispetto alla normale-> la distribuzione è PLATICURTICA

Indice di Pearson-Curtosi (II)

Se si sottrae 3 all'indice di Pearson si ottiene una versione centrata sullo zero

- Se è pari a 0 la distribuzione assume una forma NORMALE
- Se è maggiore di 0 -> la distribuzione è LEPTOCURTICA
- Se è inferiore a 0-> la distribuzione è PLATICURTICA

Esercizio 1

Consideriamo la seguente distribuzione di dati

Calcolare media, moda e mediana e commentare i risultati

Partendo dal grafico stabilire se la distribuzione è simmetrica o asimmetrica

Età	Numero di persone
20	5
23	8
27	4
30	6
40	3
65	2
totale	28

Esercizio 1-Soluzione

Calcolare media, moda e mediana

$$\mathbf{M}=[(20*5)+(23*8)+(27*4)+(30*6)+(40*3)+(65*2)]/28=(100+184+108+180+120+112)/28=\mathbf{28.71}$$

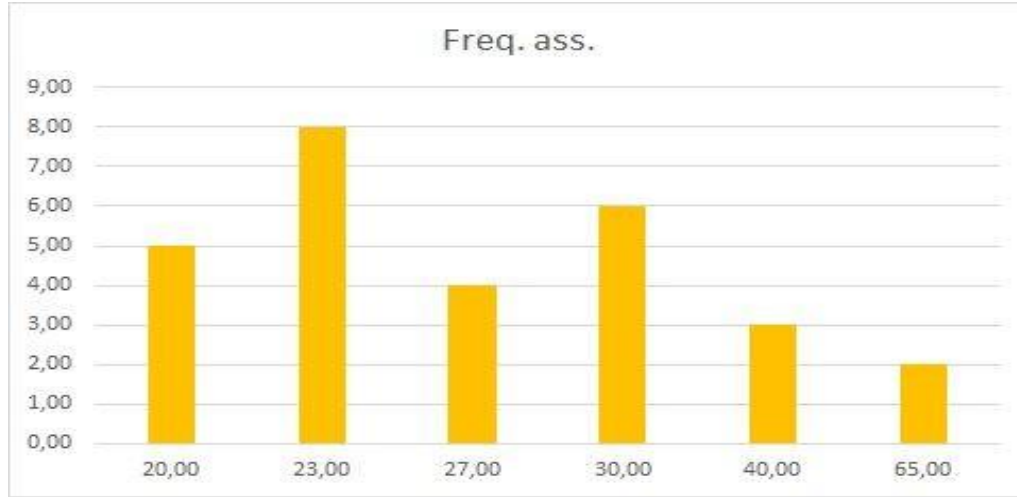
$$\mathbf{Mo= 23}$$

$$\mathbf{Me=(27+30)/2=28.5}$$

Media>Mediana>Moda

Età	Numero di persone
20	5
23	8
27	4
30	6
40	3
65	2
totale	28

Esercizio 1-Soluzione



- La distribuzione presenta un'asimmetria positive.
- La distribuzione è asimmetrica con coda a destra