

Statistica: principi e metodi



Capitolo 6 Indici di forma

Indici di forma

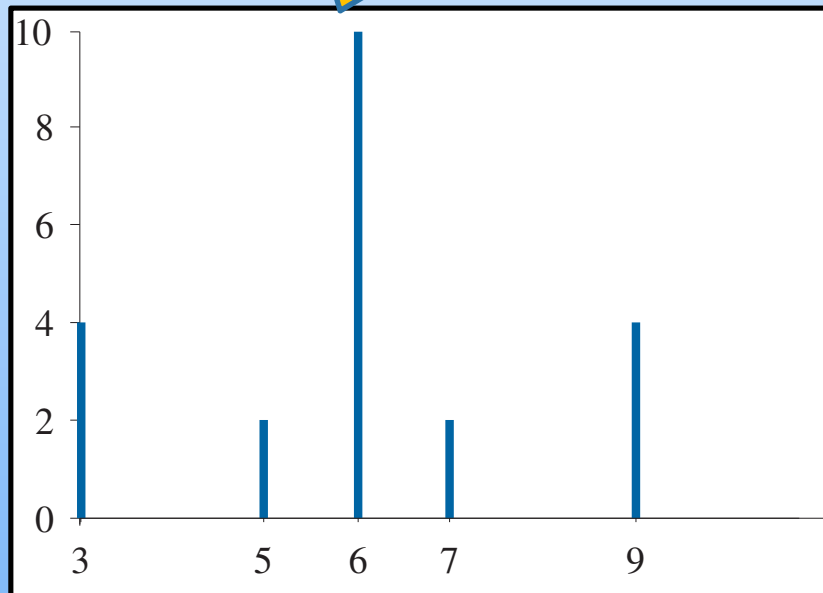
- Gli indici di forma completano il quadro degli **indicatori sintetici** delle caratteristiche delle distribuzioni statistiche di caratteri quantitativi: fin qui abbiamo visto le medie e gli indici di variabilità
- L'introduzione degli indici di forma richiede che ci si riferisca a **distribuzioni di frequenze**

Asimmetria

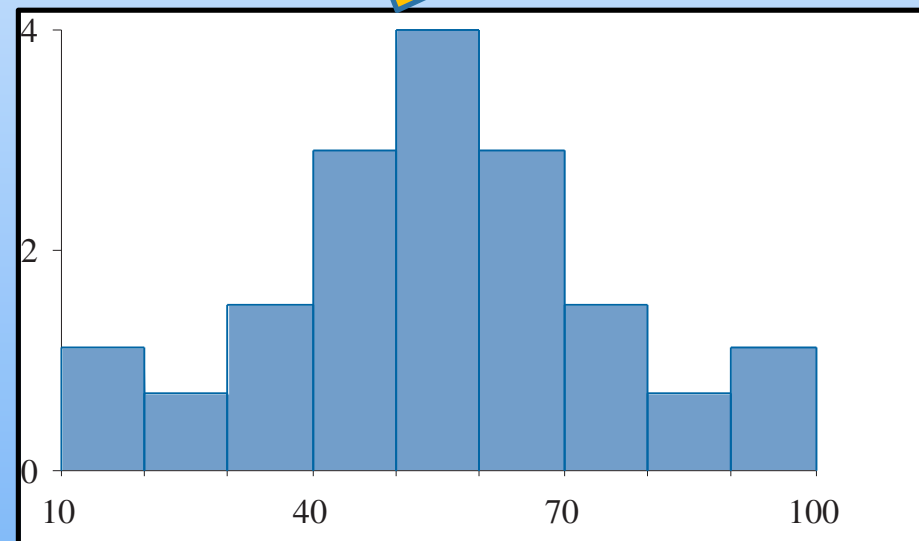
- Misura il grado di "allontanamento" della distribuzione dalla forma simmetrica
- Una distribuzione si dice simmetrica se è possibile individuare (analizzando l'istogramma) un asse verticale che tagli la distribuzione in due parti specularmente uguali

Esempi di distribuzioni simmetriche

Distribuzione di frequenze a modalità singole



Distribuzione di frequenze con modalità raggruppate in classi



Asimmetria per le distribuzioni di frequenze a modalità singole

- Considerando le seguenti coppie di modalità: la prima e l'ultima, la seconda e la penultima, la terza e la terzultima ecc., la distribuzione si dice **simmetrica** se per ciascuna delle coppie suddette **le modalità sono equidistanti dalla mediana e hanno la stessa frequenza**
- Una distribuzione non simmetrica presenterà asimmetria positiva o negativa a seconda che siano più frequenti nella distribuzione le modalità più piccole o più grandi

Proprietà di una distribuzione simmetrica

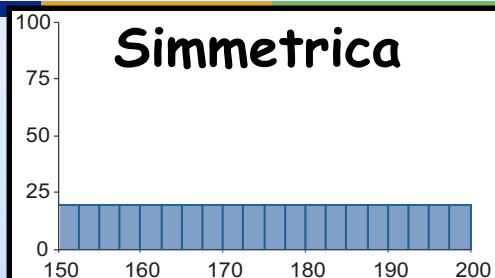
- La media aritmetica coincide con la mediana
- La somma degli scarti dalla media aritmetica elevati a una potenza dispari è uguale a 0
- Il primo e terzo quartile hanno la stessa distanza dalla mediana

Asimmetria positiva e asimmetria negativa

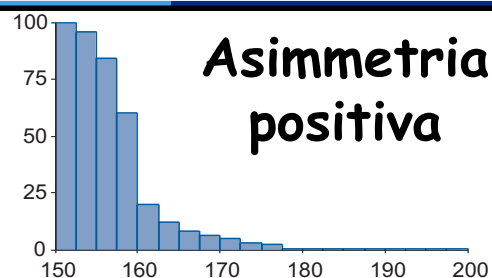
Preso come riferimento il punto dell'asse delle ascisse che rappresenta la **mediana** della distribuzione, si ha

- **asimmetria positiva** se il grafico tende a "prolungarsi" maggiormente **a destra di tale punto**;
- **asimmetria negativa** se il prolungamento è **a sinistra**

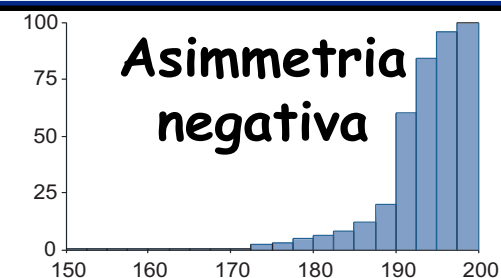
Asimmetria positiva e asimmetria negativa



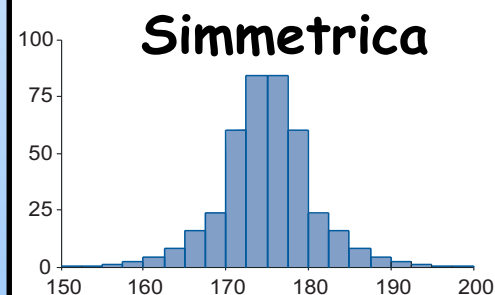
(a)



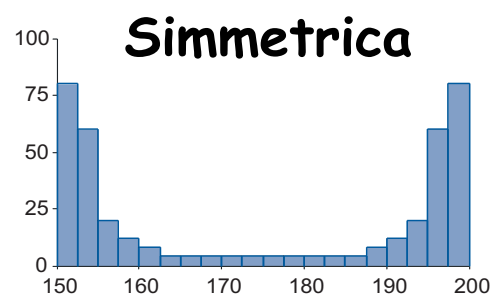
(b)



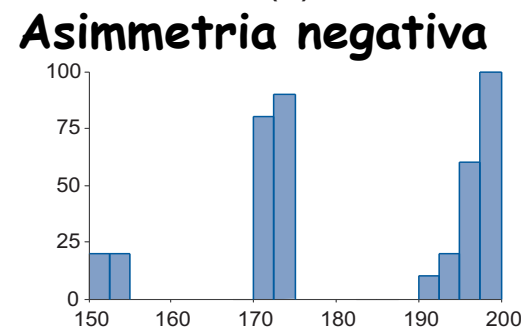
(c)



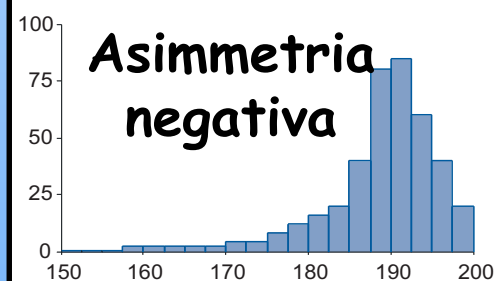
(d)



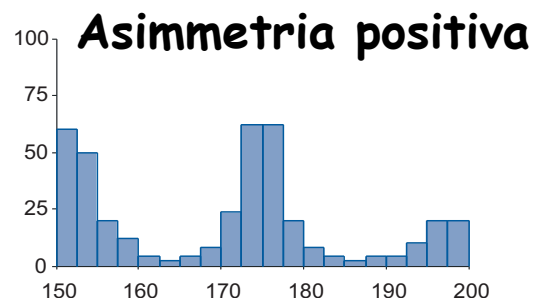
(e)



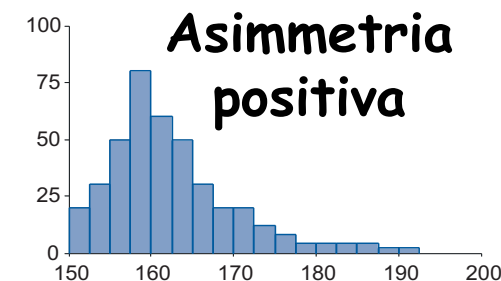
(f)



(g)



(h)



(i)

Indici di asimmetria

- Una misura dell'asimmetria è data da

$$a_1 = \frac{1}{\sigma^3} \left[\frac{1}{N} \sum_{i=1}^k (x_i - \mu)^3 \cdot n_i \right]$$

- Una seconda misura è data da

$$a_2 = \frac{(q_3 - m) - (m - q_1)}{(q_3 - m) + (m - q_1)}$$

m , q_1 e q_3 sono la mediana, il primo e il terzo quartile.

- *Se $a_1 > 0$, $a_2 > 0$ si ha asimmetria positiva, se $a_1 < 0$, $a_2 < 0$ si ha asimmetria negativa*
- *Se $a_2 = 0$ non è detto che la distribuzione sia simmetrica*

Indici α_1 e α_2 : esempio

$$\alpha_1 = \frac{1}{\sigma^3} \left[\frac{1}{N} \sum_{i=1}^k (x_i - \mu)^3 \cdot n_i \right]$$

$$\alpha_2 = \frac{(q_3 - m) - (m - q_1)}{(q_3 - m) + (m - q_1)}$$

Distribuzione degli esami sostenuti
da un laureando

Voto	N. esami	$(x_i - \mu)^2 \cdot n_i$	$(x_i - \mu)^3 \cdot n_i$	N_i
23	1	14.19	-53.44	1
24	3	22.96	-63.53	4
25	4	12.48	-22.06	8
26	7	4.11	-3.15	15
27	5	0.27	0.06	20
28	4	6.08	7.50	24
29	1	4.99	11.14	25
30	5	52.27	169.01	30
Totale	30	117.35	45.53	

□ La media è $\mu = 26.77$

$$\alpha_1 = \frac{45.53 / 30}{(\sqrt{117.35 / 30})^3} = 0,196$$

$$q_1 = 25 \quad m = 26.5 \quad q_3 = 28$$

$$\alpha_2 = \frac{(28 - 26.5) - (26.5 - 25)}{(28 - 26.5) + (26.5 - 25)} = 0$$

Curtosi

È il grado di "allontanamento" della distribuzione di frequenze effettiva dal **modello teorico** rappresentato dalla **distribuzione normale**

Distribuzione normale

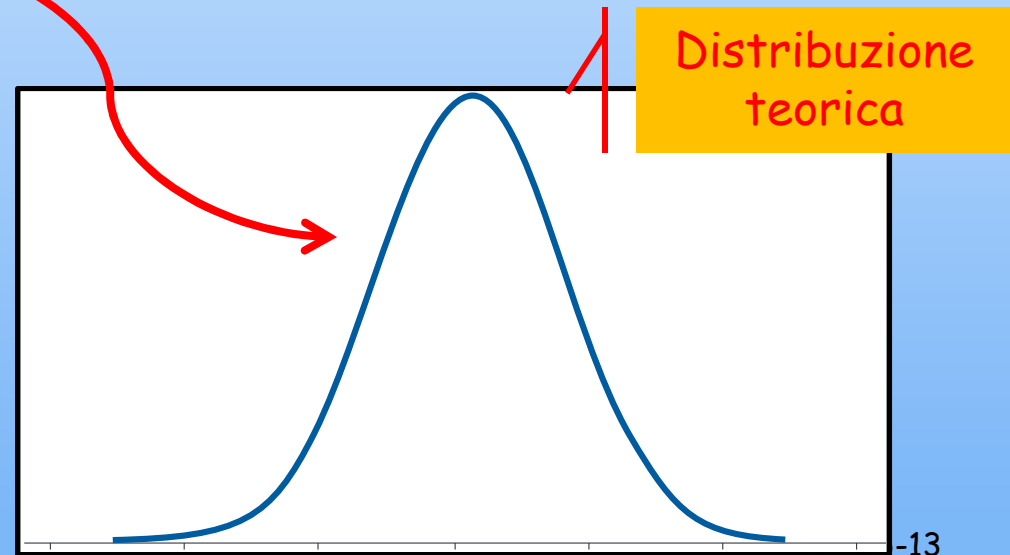
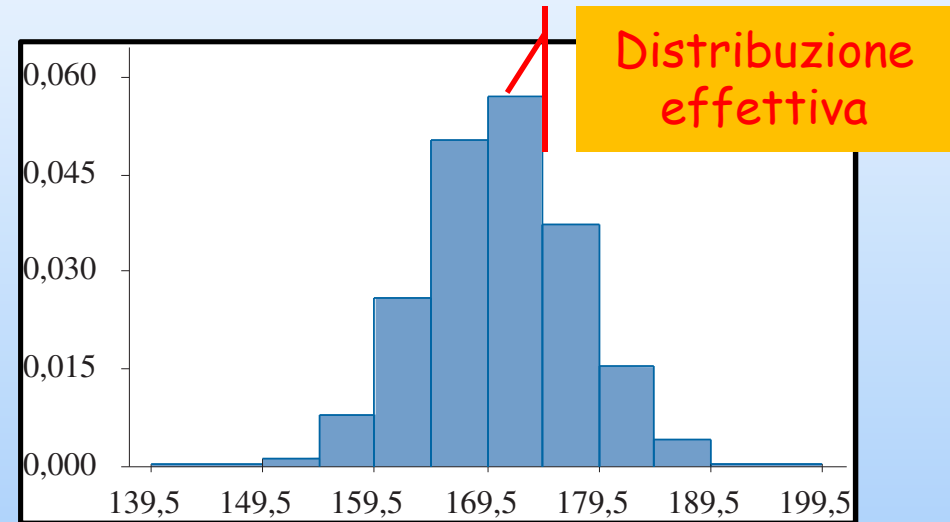
- È un modello matematico che può essere impiegato per approssimare una distribuzione di frequenze reale
- Funzione di densità normale (o curva di Gauss) è

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}$$

- x assume tutti i valori dell'asse reale
- μ e σ sono due costanti numeriche interpretabili come media e deviazione standard della distribuzione teorica

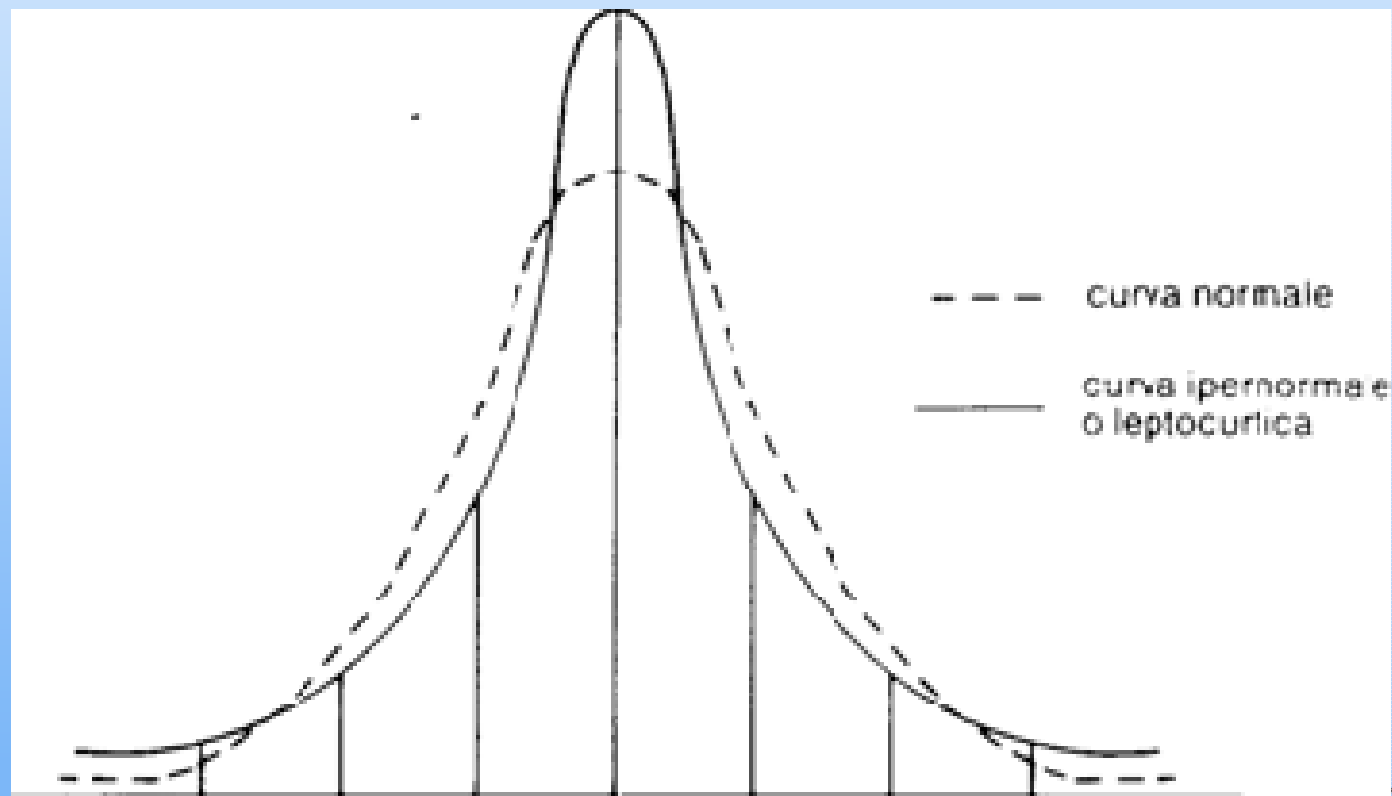
Distribuzione normale

- La media μ coincide con il punto di **massima densità**
- σ determina il **grado di concentrazione** della curva attorno alla media:
 - ✓ a valori crescenti di σ corrispondono curve sempre più schiacciate verso l'asse delle ascisse.
- L'**area** sottesa alla curva è uguale a 1



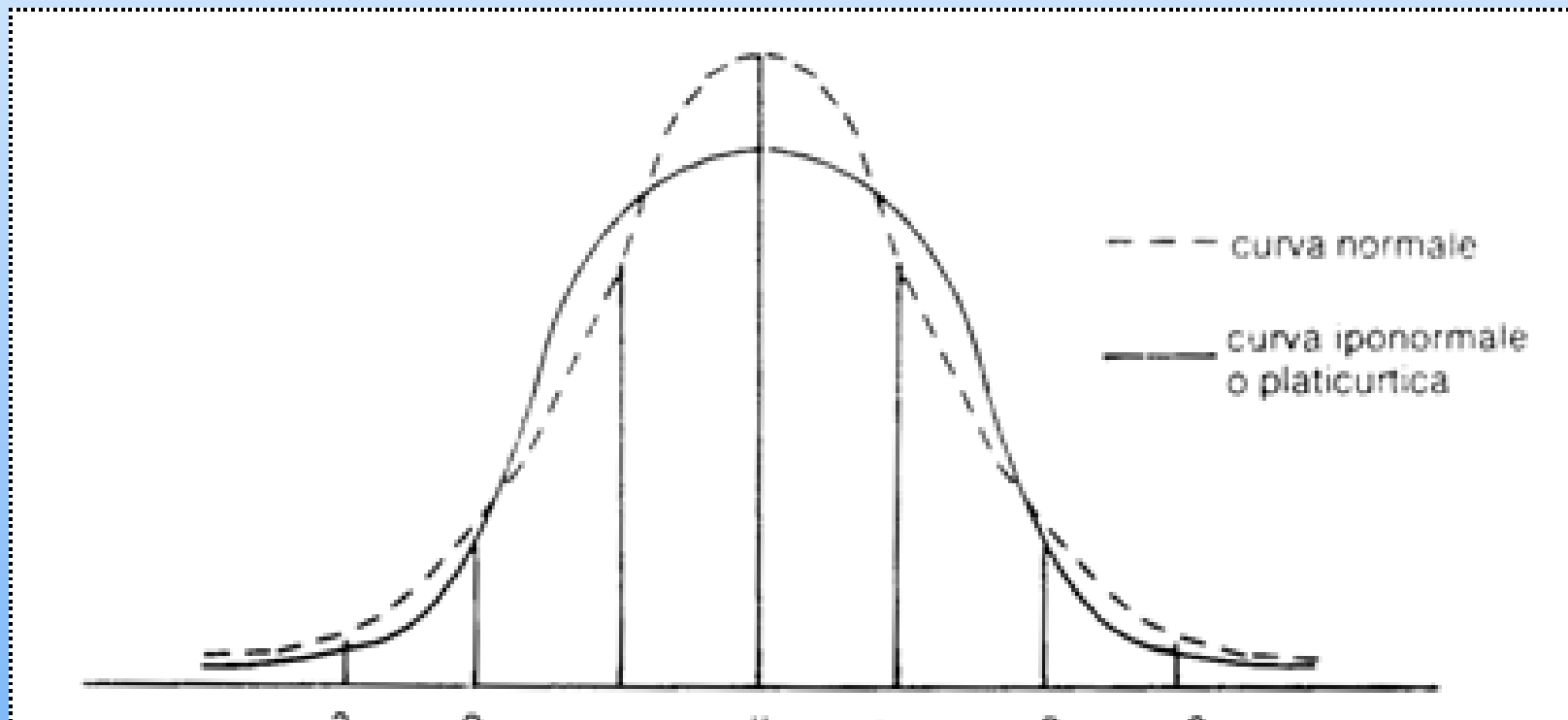
Ipernormalità

La **ipernormalità (curva leptocurtica)** è caratterizzata da ordinate più elevate rispetto alla normale attorno al valore medio e alle code.



Ipernormalità e iponormalità

La **iponormalità** (curva platicurtica) è caratterizzata da ordinate meno elevate attorno al valore medio e alle code



Indice di curtosi

- Una misura della curtosi è data

$$\nu = \frac{1}{\sigma^4} \left[\frac{1}{N} \sum_{i=1}^k (x_i - \mu)^4 \cdot n_i \right] - 3$$

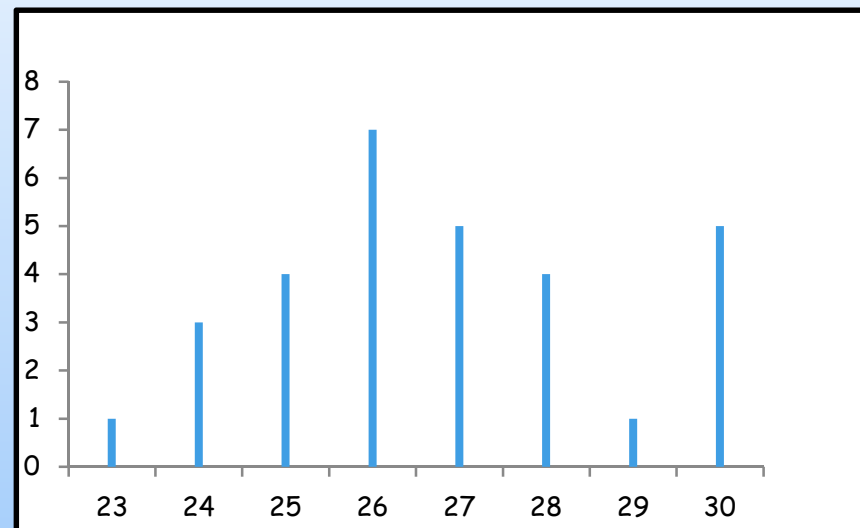
- Assume valori vicini allo **0** se la distribuzione è molto prossima alla normale
- È **positivo** nel caso di ipernormalità;
- È **negativo** nel caso di iponormalità.

Indice di curtosi: calcolo

$$\gamma = \frac{1}{\sigma^4} \left[\frac{1}{N} \sum_{i=1}^k (x_i - \mu)^4 \cdot n_i \right] - 3$$

Esami sostenuti da un laureando

Voto	N. esami	$(x - \mu)^4 \cdot n_i$
23	1	201.29
24	3	175.77
25	4	38.97
26	7	2.42
27	5	0.01
28	4	9.26
29	1	24.88
30	5	546.48
Totale	30	999.08



$$\mu = 26.77; \sigma = 1.98$$

□ Indice di curtosi

$$\gamma = \frac{1}{1.98^4} \left[\frac{1}{30} \cdot 999.08 \right] - 3 = -0.82$$

□ La distribuzione è iponormale