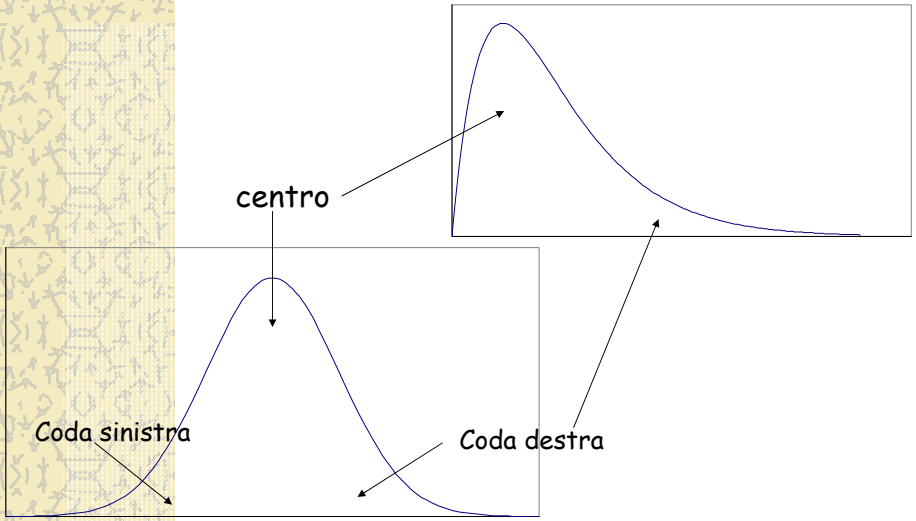
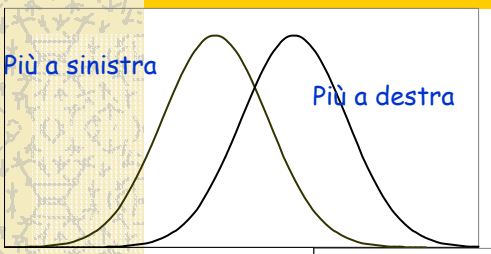
Aspetti notevoli delle distribuzioni

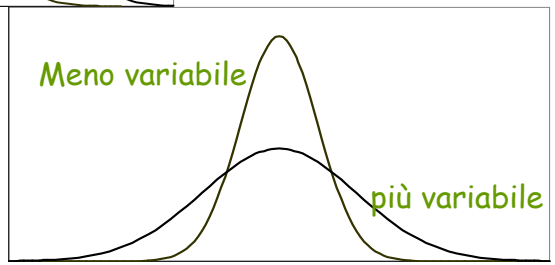


Aspetti caratterizzanti le distribuzioni: posizione e variabilità

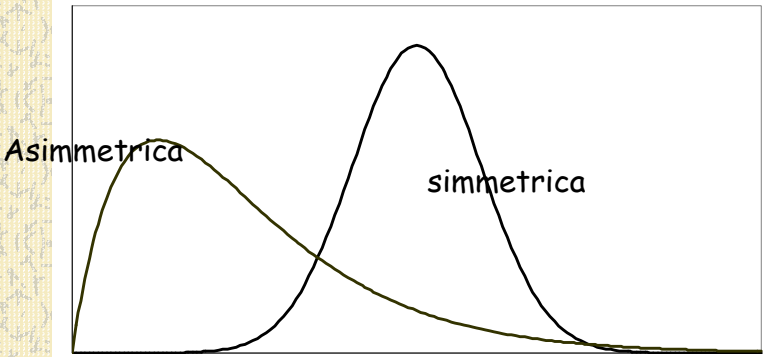


Posizione

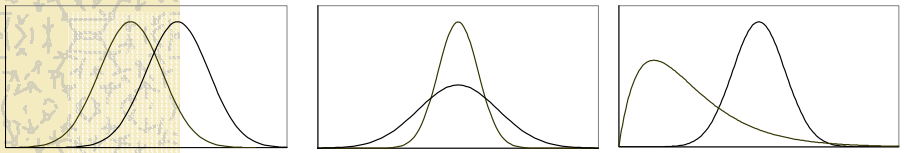
Variabilità

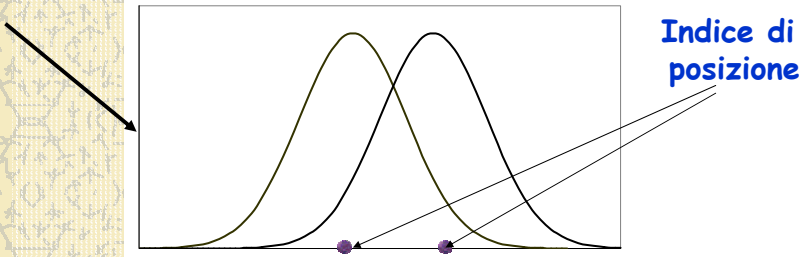


Aspetti caratterizzanti le distribuzioni: forma



Indici caratteristici delle distribuzioni





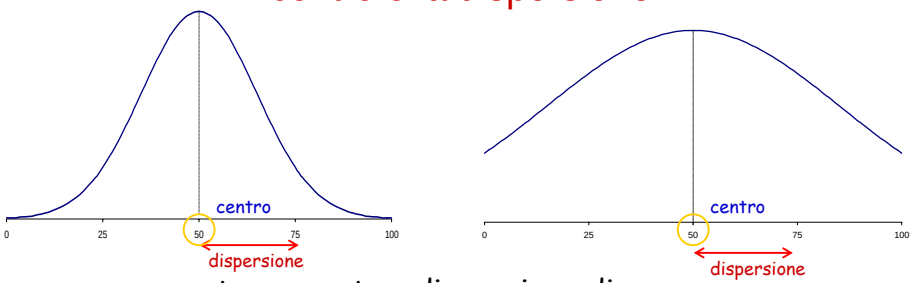
Gli indici di posizione: medie

Sintesi della distribuzione attraverso un valore rappresentativo

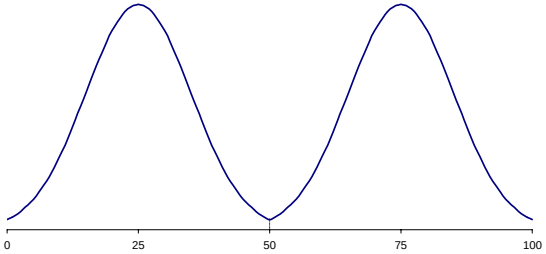
VARIABILI

- sconnesse: moda
- ordinali: mediana
- quantitative: media aritmetica, media geometrica

Il centro e la dispersione



stesso centro, dispersione diversa



Il centro non sintetizza la distribuzione!

Media aritmetica

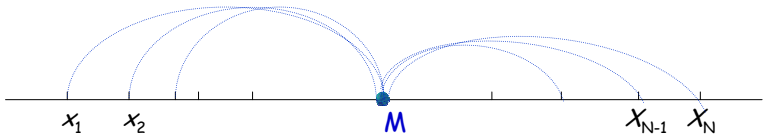
X v.s. quantitativa

$\{x_1, x_2, \ldots, x_N\}$ successione

$$\mu = M = \frac{1}{N}\sum_{i=1}^{N} x_i$$

esempio X: 2 3 2 1 ➔ M=(2+3+2+1)/4=2

MEDIA: CENTRO DELL'INSIEME DEGLI *N* PUNTI



Distanza tra x_i e x_j in $\mathbb{R}$: $d_{ij} = |x_i - x_j|$

Media aritmetica di una serie

Tabella di frequenza

Mod.tà	Freq.	Fr.rel.
x_1	n_1	f_1
x_2	n_2	f_2
...	...	...
x_j	n_j	f_j
...	...	...
x_J	n_J	f_J
Totale	N	1

$$M = \frac{1}{N}\sum_{j=1}^{J} x_j n_j = \sum_{j=1}^{J} x_j f_j$$

Serie: $\{(x_j, n_j)\}$

j=1,2,…,J

Esempio media aritmetica di una serie

ETA	nj	fj*100	Fj	Fj*100
20	54	53.47	54	53.47
21	21	20.79	75	74.26
22	17	16.83	92	91.09
23	3	2.97	95	94.06
25	1	0.99	96	95.05
26	1	0.99	97	96.04
29	1	0.99	98	97.03
39	1	0.99	99	98.02
40	1	0.99	100	99.01
46	1	0.99	101	100.00

Età media=21

Media aritmetica per dati in classi

Ipotesi istogramma:

equidistribuzione frequenze all'interno delle classi

Tabella di frequenza

Mod.tà	Freq.	Fr.rel.
x_0-x_1	n_1	f_1
x_1-x_2	n_2	f_2
...	...	...
x_{j-1}-x_j	n_j	f_j
...	...	...
x_{J-1}-x_J	n_J	f_J
Totale	N	1

Valore centrale di classe

$c_j = (x_j + x_{j-1})/2$

$$M = \frac{1}{N}\sum_{j=1}^{J} c_j n_j = \sum_{j=1}^{J} c_j f_j$$

Seriazione: $\{(x_{j-1}; x_j), n_j)\}$

j=1,2,...,J

Esempio: media di una seriazione

CFU	nj	fj*100	Fj	Fj*100
0-24	14	15.91	14	15.91
25-40	34	38.64	48	54.55
41-50	19	21.59	67	76.14
>50	21	23.86	88	100.00

Frequency Missing = 14
MIN=9, MAX=142, P90%=69

foglio di calcolo

Proprietà della media aritmetica

- Centro di ordine 2 (minimi quadrati)
- Lascia invariato l'ammontare complessivo: $NM = \sum_{i=1}^{N} x_i$
- Baricentro $\sum_{i=1}^{N} (x_i - M) = 0$
- Internalità (propr. di Cauchy) $x_{[1]} \le M \le x_{[n]}$
- Invarianza per trasformazioni lineari

$$Y = a + bX \Rightarrow M(Y) = a + bM(X)$$

Media aritmetica ponderata

$$M_p = \frac{\sum_{j=1}^{J} x_j p_j}{\sum_{j=1}^{J} p_j}$$

- Esempio: indici sintetici dei prezzi con pesi pari al valore delle merci scambiate

  (A) Indice di LASPEYRES: pesi= $p_{iA}q_{iA}$

  (B) Indice di PAASCHE: pesi= $p_{iA}q_{iB}$

  TEMPI: A<B

Indici sintetici dei prezzi

- Paniere di beni: i=1,…,k
- NI elementari $I_i=p_{iB}/p_{iA}$ A<B
- Indice di Laspeyres

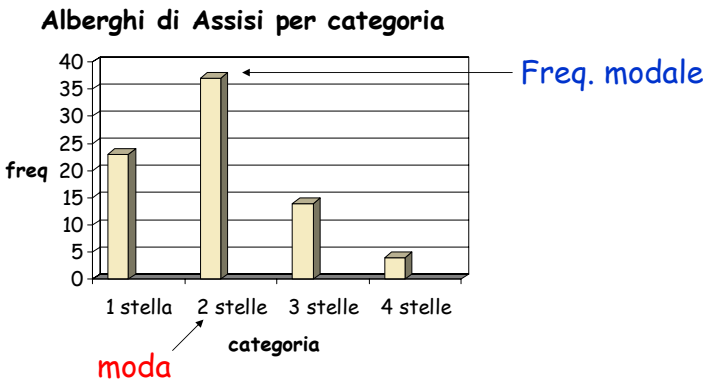
$$I_L = \frac{\sum_{i=1}^{k} \frac{p_{iB}}{p_{iA}} \times p_{iA}q_{iA}}{\sum_{i=1}^{k} p_{iA}q_{iA}} = \frac{\sum_{i=1}^{k} p_{iB}q_{iA}}{\sum_{i=1}^{k} p_{iA}q_{iA}}$$

- Indice di Paasche

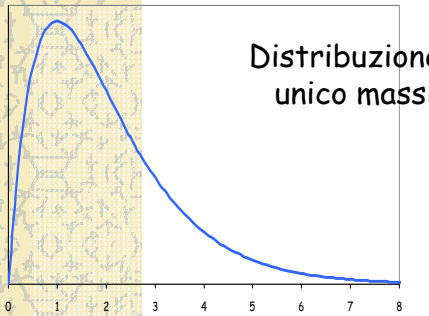
$$I_L = \frac{\sum_{i=1}^{k} \frac{p_{iB}}{p_{iA}} \times p_{iA}q_{iB}}{\sum_{i=1}^{k} p_{iA}q_{iB}} = \frac{\sum_{i=1}^{k} p_{iB}q_{iB}}{\sum_{i=1}^{k} p_{iA}q_{iB}}$$

La moda

Modalità cui corrisponde la frequenza più alta.

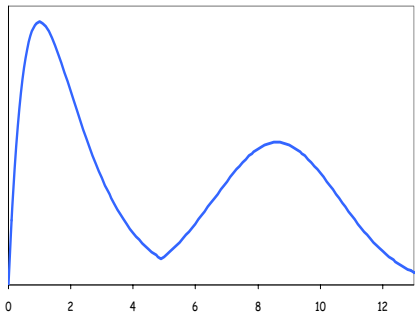


Moda
utile per distribuzioni unimodali



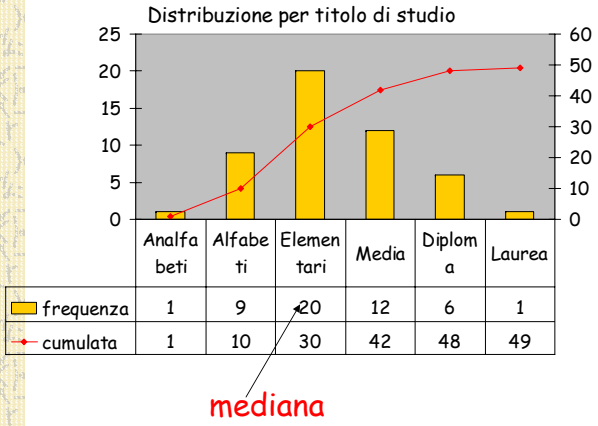
Distribuzione unimodale
unico massimo locale

Distribuzione bimodale
più di un massimo locale



La mediana

Modalità che è più grande del 50% delle osservazioni e più piccola del 50% delle osservazioni



Mediana di una successione

La mediana M_e di *n* numeri ordinati in senso non decrescente $\{x_{(1)}, \ldots, x_{(n)}\}$ è:

- per *n* dispari $M_e = x_{(n+1)/2}$
- per *n* pari M_e in $[x_{n/2}; x_{(n/2)+1}]$

se X è quantitativa, si considera

$$M_e = [x_{n/2} + x_{(n/2)+1}]/2$$

Calcolo della mediana attraverso le frequenze cumulate

(1) Calcolare **0.5*N**=N/2

(2) guardare il valore ottenuto:

- se *N*0.5* è un numero frazionario (N dispari) e *M* numero intero tale che *M<N*0.5<M+1:*

$$M_e = x_{(M+1)}$$

- se *N*0.5 è* numero intero (N pari):

$$M_e \in [x_{(N*0.5)}; x_{(N*0.5+1)}]$$

(3) Individuare il valore mediano in base al valore delle frequenze cumulate

Calcolo della mediana attraverso la cumulata relativa F(X)

$$M_e = x_m : \begin{cases} F(x_m) \geq 0.5 \\ F(x_{m-1}) \leq 0.5 \end{cases}$$

Se x è una v.s. ordinale o discreta si ha:

$$F(x_{m-1}) = F(x_m) - f_m$$

Proprietà della mediana

- Centro di ordine 1 $\quad M_e : \sum_{i=1}^{n} |x_i - M_e| = \min$

- Si può utilizzare anche per v.s. ordinali

- M_e resta invariata se si sostituiscono i termini x< M_e o x> M_e : non risente di valori anomali

Calcolo della mediana per v.s. ordinali

$$M_e = x_{0.5} : \begin{cases} pr(X \geq x_{0.5}) \geq 0.5 \\ pr(X \leq x_{0.5}) = F(x_{0.5}) \geq 0.5 \end{cases}$$

TITSTUD	Freq	Percent	Cumulative Frequency	Cumulative Percent
lic. elementare	2	0.34	2	0.34
lic. media	105	17.89	107	18.23
diploma	413	70.36	520	**88.59**
diploma univ.	7	1.19	527	89.78
laurea	61	10.22	588	100.00

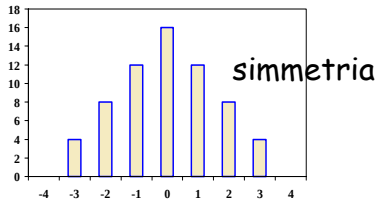
N*0.5=588*0.5=294➔ $M_e = x_{0.5} \in [x_{294}; x_{295}] = diploma$

Calcolo della mediana per v.s. ordinali

Titolo di studio	*Frequenze*	*Frequenze cumulate*
Nessun titolo	1862	1862
Licenza elementare	9903	11765
Licenza media	4491	16256
Diploma	3093	19349
Laurea	1160	20509
Totale	20509	

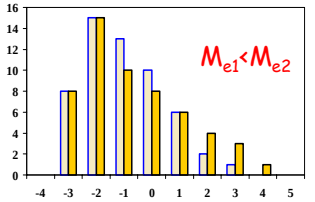
Mediana di un istogramma:

$$M_e : pr(X \leq M_e) = F(M_e) = 0.5$$

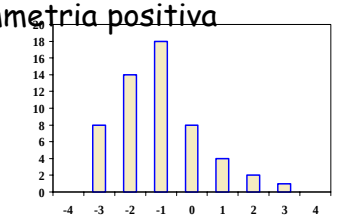


Me=M

Non risente delle code

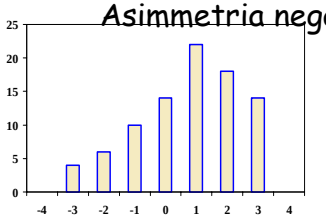


Asimmetria positiva



Me<M

Asimmetria negativa



M< Me