

Analisi delle distribuzioni univariate

Breve riepilogo pensando a una distribuzione di frequenza per caratteri continui

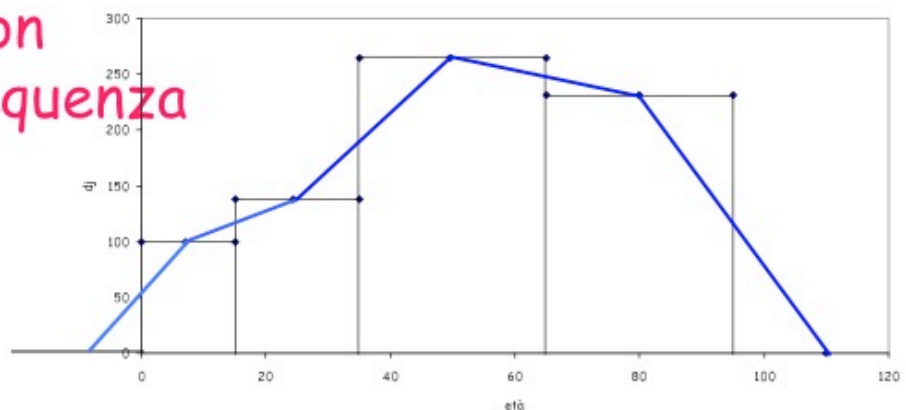
rappresentazione grafica

Tab. 4 - Consumatori di farmaci per classe di età, 1998 (migliaia).
Istat, Rapporto sull'Italia 2001

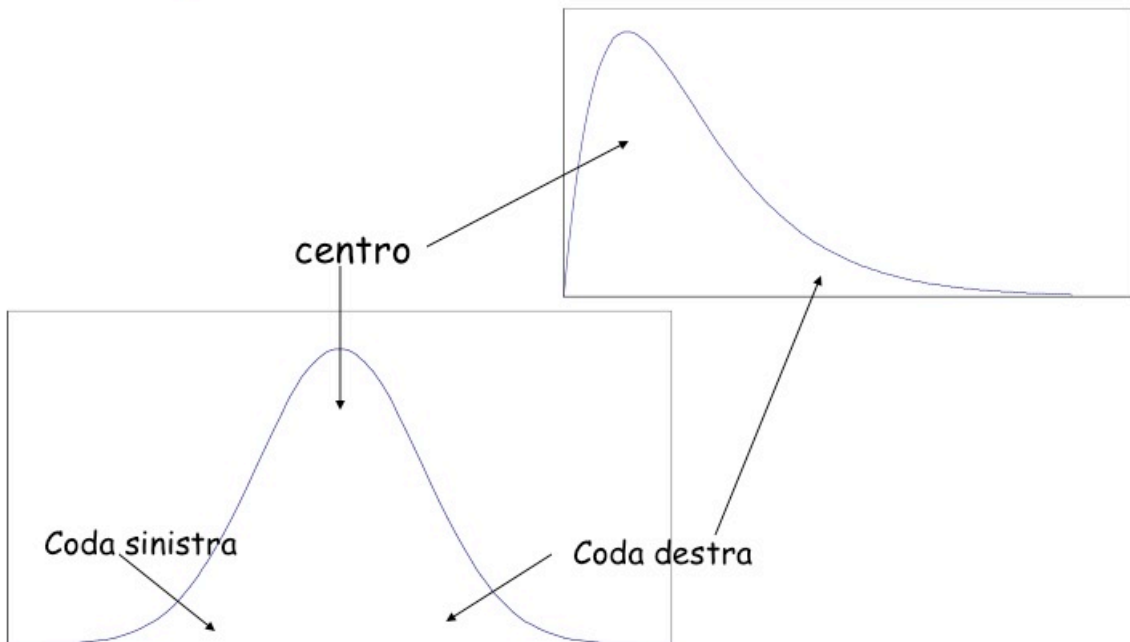
età	consumator	aj	dj
0-14	1396	14	99.71429
15-34	2626	19	138.2105
35-64	7687	29	265.069
65 e oltre	6933	30	231.1
TOTALE	18642		

Poligono di frequenza: segmenti che uniscono i punti centrali dei lati superiori dei rettangoli che definiscono l'istogramma, comprese due classi terminali con frequenza zero e ampiezza pari all'ampiezza della classe adiacente

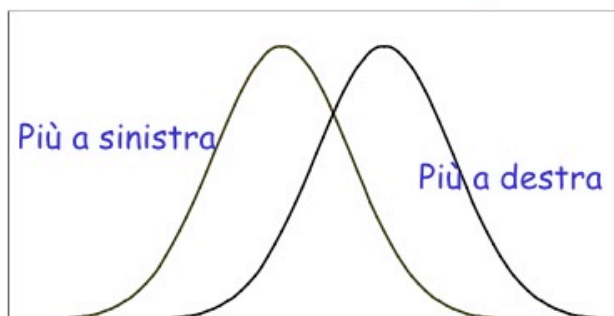
Istogramma con poligono di frequenza



Aspetti notevoli delle distribuzioni

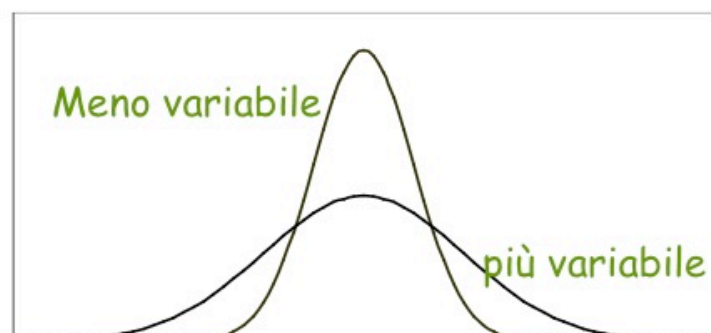


Aspetti caratterizzanti le distribuzioni: posizione e variabilità



Posizione

Variabilità



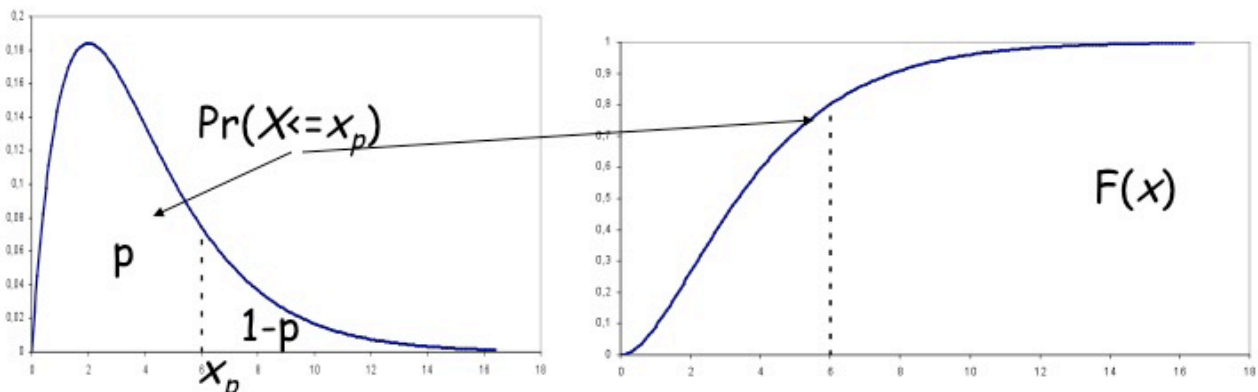
Gli indici di posizione: medie

Sintesi della distribuzione attraverso un valore rappresentativo

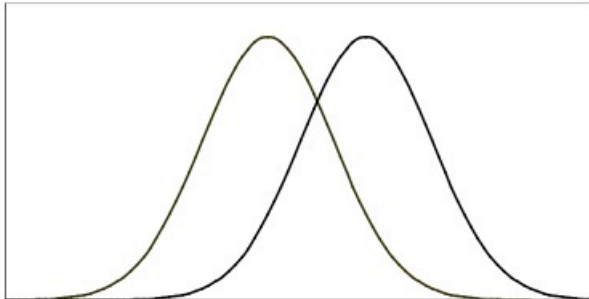
- Distribuzioni sconnesse: moda
- Distribuzioni ordinate: mediana
- Seriazioni: media aritmetica
... e altre medie

Indici di posizione notevoli: i quantili

- $p=0.01,0.02,\dots,0.98,0.99$ Percentili
- $p=0.1,0.2,\dots,0.8,0.9$ Decili
- $p=0.25,0.50,0.75$ Quartili
- $p=0.5$ Mediana

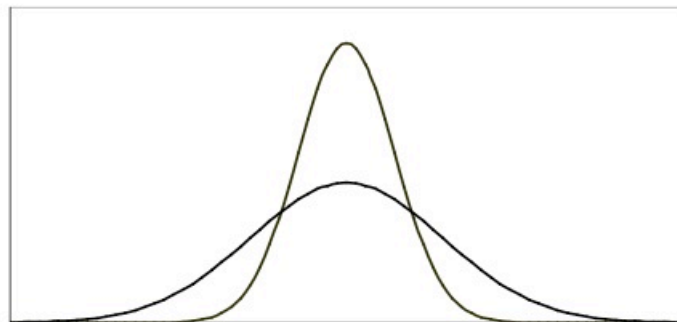


VARIABILITÀ



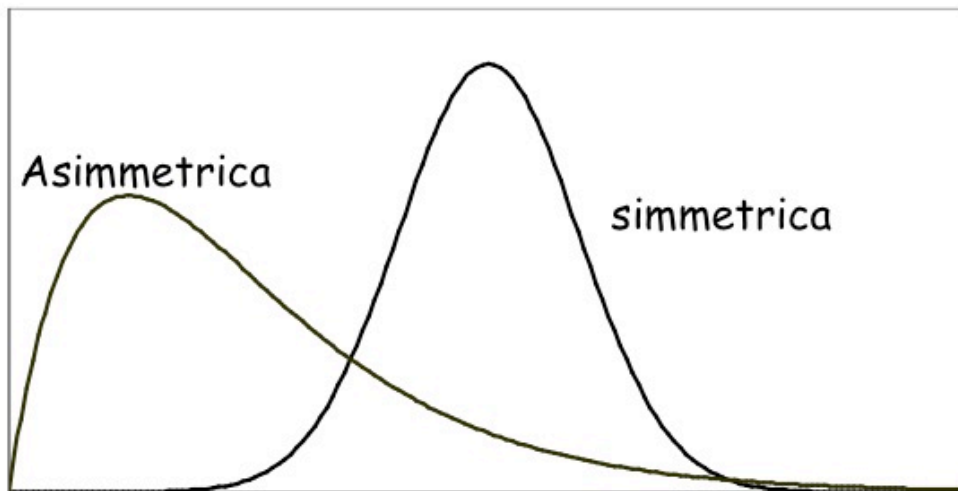
media diversa,
stessa variabilità

stessa media,
Variabilità diversa



	Assoluti	Relativi
basati sullo scostamento dalla media	Devianza $\rightarrow \sum (x_j - \bar{x})^2$ Varianza $\rightarrow \sigma^2 = \frac{1}{N} \sum (x_j - \bar{x})^2$ Scost. quadratico medio $\sigma = \sqrt{\frac{1}{N} \sum (x_j - \bar{x})^2}$	Coefficiente di variazione $CV = \frac{\sigma}{\bar{x}}$
basati sulle differenze tra modalità	Differenza media $\Delta = \frac{\sum_i \sum_j x_i - x_j }{N(N-1)}$	Rapporto di concentrazione $R = \frac{\Delta}{2\bar{x}}$

Aspetti caratterizzanti le distribuzioni: forma



il confronto "media → mediana" dà indicazioni sulla simmetria/asimmetria