

Elementi di Statistica descrittiva

Gli Indici di VARIABILITA'

- **Campo di variazione**
- **Scarto dalla media**
- **Varianza**
- **Scarto quadratico medio**
- **Coefficiente di variazione**

Indici di Variabilità

I valori medi sono indici importanti per la descrizione sintetica di un fenomeno statistico

Hanno però il limite di non darci alcuna informazione sulla distribuzione dei dati

In statistica è possibile valutare in modo sintetico
la distribuzione dei dati mediante gli indici di
variabilità (o dispersione)

Vedremo i seguenti indici

- Campo di variazione (Range)
- Scarto medio dalla media
- Varianza e scarto quadratico medio
- Coefficiente di variazione

Campo di variazione

E' il più semplice degli indici di variazione:

Si calcola facendo la differenza tra il dato più grande e il dato più piccolo

$$\text{Campo variazione} = X_{\max} - X_{\min}$$

Rappresenta l'ampiezza dell'intervallo dei dati

Osservazioni:

1. Il campo di variazione dà informazioni sulla distribuzione dei dati:
 - più R è piccolo più i dati sono concentrati;
 - più R è grande più i dati sono dispersi.
2. R è espresso nella stessa unità di misura dei dati
3. Tuttavia R tiene conto solo dei dati estremi della distribuzione e non di tutti i dati, pertanto distribuzioni diverse ma con gli stessi valori estremi hanno range uguali

Scarto medio dalla media aritmetica

Un altro modo per calcolare la variabilità dei dati (tenendo conto di tutti i dati) consiste nel calcolare la distanza di tutti i dati dalla media e fare la media aritmetica di tali distanze

$$\text{Scarto medio} = S_m = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

Scarto medio = Distanza media dei dati dalla media

Osservazioni:

1. Lo scarto medio dalla media dà informazioni sulla distribuzione dei dati:
 - più $\underline{S_M}$ è piccolo più i dati sono concentrati;
 - più $\underline{S_M}$ è grande più i dati sono dispersi.
2. $\underline{S_M}$ è espresso nella stessa unità di misura dei dati
3. Non ha l'inconveniente del “Campo di variazione” in quanto $\underline{S_M}$ tiene conto di tutti i dati della distribuzione

Varianza e Scarto quadratico medio

Sono gli indici di variabilità più utilizzati, e tengono conto della distribuzione di tutti i dati.

Varianza

Rappresenta la media aritmetica dei quadrati delle distanze dei dati dalla media \bar{x}

$$\text{Varianza} = \sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

$$\text{Varianza} = \frac{\sum_1^n (x_i - \bar{x})^2}{n} = \frac{\sum_1^n (\Delta x_i)^2}{n}$$

Scarto quadratico medio o

Deviazione standard

È uguale alla radice quadrata della varianza

$$\text{Scarto quadr. medio} = \sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

$$\text{Scarto quadr medio} = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_1^n (\Delta x_i)^2}{n}}$$

Osservazioni:

1. La varianza σ^2 e lo scarto quadratico medio σ danno informazioni sulla distribuzione dei dati:
 - più σ^2 e σ sono piccoli più i dati sono concentrati;
 - più σ^2 e σ sono grandi più i dati sono dispersi.
2. Entrambi gli indici tengono conto di tutti i dati della distribuzione

3. Entrambi si basano sulla proprietà della media per cui la somma dei quadrati degli scarti dalla media è minima
4. La varianza è espressa mediante il quadrato dell'unità di misura dei dati
5. Lo scarto quadratico nella stessa unità di misura dei dati e pertanto viene preferito alla varianza

Il coefficiente di variazione CV

Il CV è una misura relativa di dispersione (le precedenti sono misure assolute) ed è una grandezza adimensionale.

E' particolarmente utile quando si devono confrontare le distribuzioni di due gruppi con medie molto diverse o con dati espressi in scale differenti (es. confronto tra variazione del peso e variazione dell'altezza).

$$CV = \left(\frac{\sigma}{x} \cdot 100 \right) \%$$

In natura il coeff. di variazione tende a rimanere costante per ogni fenomeno:
i valori normalmente variano dal 5% al 15%

Se i valori di CV sono esterni a quelli indicati o si è in presenza di errori di rilevazione, oppure il fenomeno presenta aspetti particolari.

Le misure di Forma

Sono indici sintetici utilizzati per evidenziare particolarità nella forma della distribuzione.

Noi esamineremo:

- l'asimmetria
- la curtosi

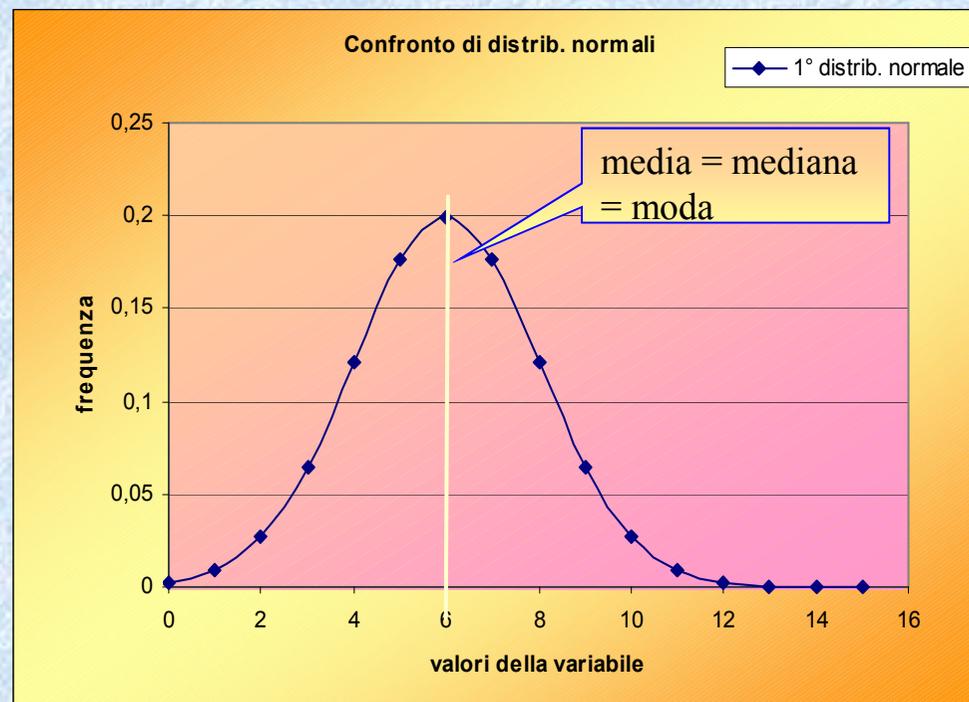
Asimmetria

Una distribuzione è simmetrica quando la sua curva di frequenza presenta un asse di simmetria

In una distribuzione simmetrica media, mediana e moda sono coincidenti.

In una distribuzione asimmetrica media, mediana e moda non sono più coincidenti

e proprio la differenza (distanza) tra la media e la moda può essere considerata una misura della asimmetria



Sono state proposte diverse misure dell' asimmetria, per esempio le più semplici sono:

$$asimmetria = \frac{\text{media aritmetica} - \text{moda}}{\text{scarto quadratico medio}} = \frac{\bar{x} - \text{moda}}{\sigma}$$

$$asimmetria = \frac{3(\text{media aritmetica} - \text{mediana})}{\text{scarto quadratico medio}} = \frac{3(\bar{x} - \text{mediana})}{\sigma}$$

Dette rispettivamente: primo e secondo coeff. di asimmetria di Pearson

σ = scarto quadratico medio

Se $a = 0$ distribuzione simmetrica

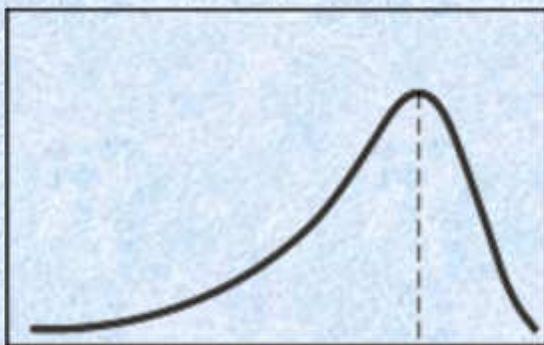
Se $a > 0$ asimmetria destra

Se $a < 0$ asimmetria sinistra

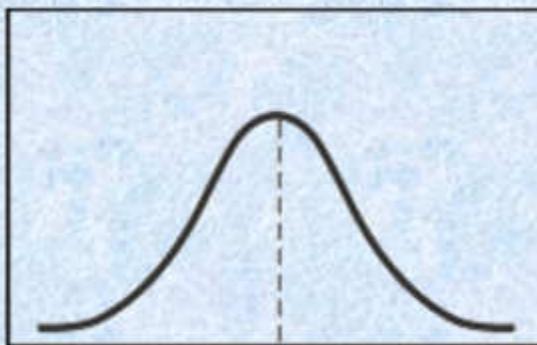
Forma della distribuzione: Simmetrica o Obliqua

Per descrivere la forma della distribuzione è sufficiente confrontare la media con la mediana. Se queste due misure sono uguali, la distribuzione è considerata simmetrica.

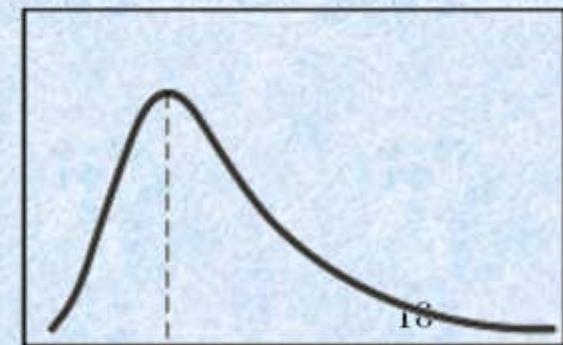
- media < mediana:
asimmetria negativa o distribuzione obliqua a sinistra
- media = mediana: **simmetria**
- media > mediana:
asimmetria positiva o distribuzione obliqua a destra



(a) Obliqua a sinistra



(b) Simmetrica



(c) Obliqua a destra

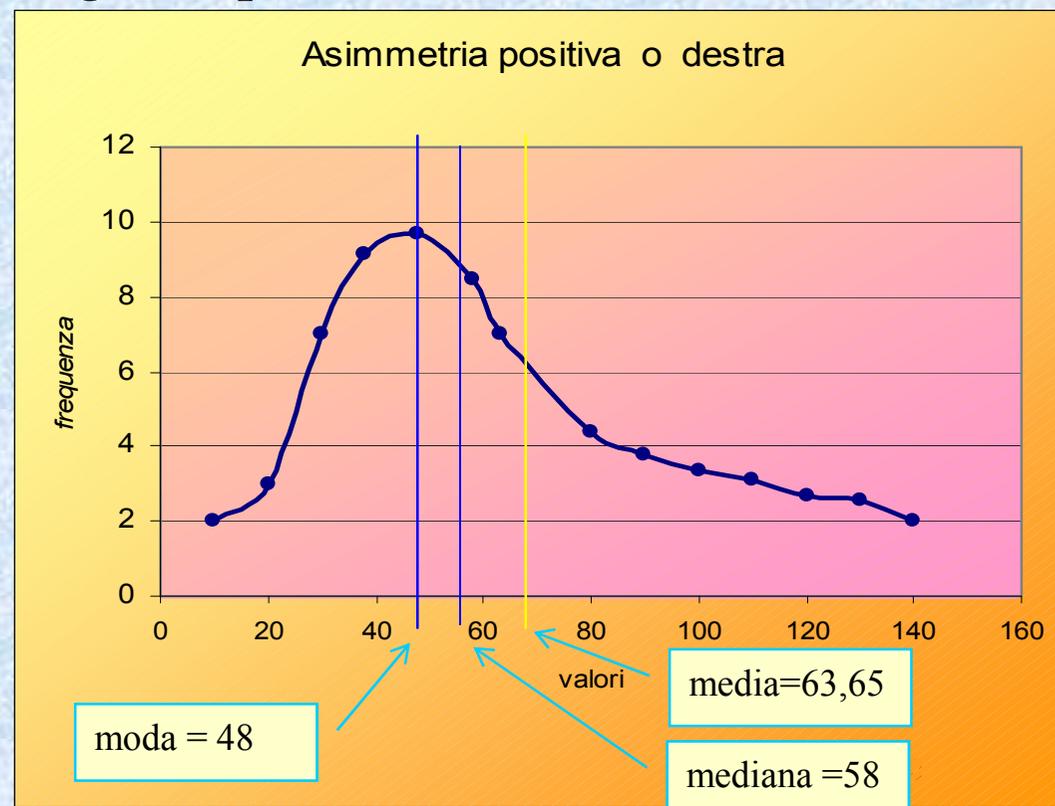
Asimmetria positiva (as. Destra)

La distribuzione è asimmetrica quando non presenta nessun asse di simmetria.

Si ha un'asimmetria positiva o destra quando il ramo destro della curva è più lungo di quello sinistro

In questo caso si ha:

$$\text{moda} < \text{mediana} < \text{media}$$

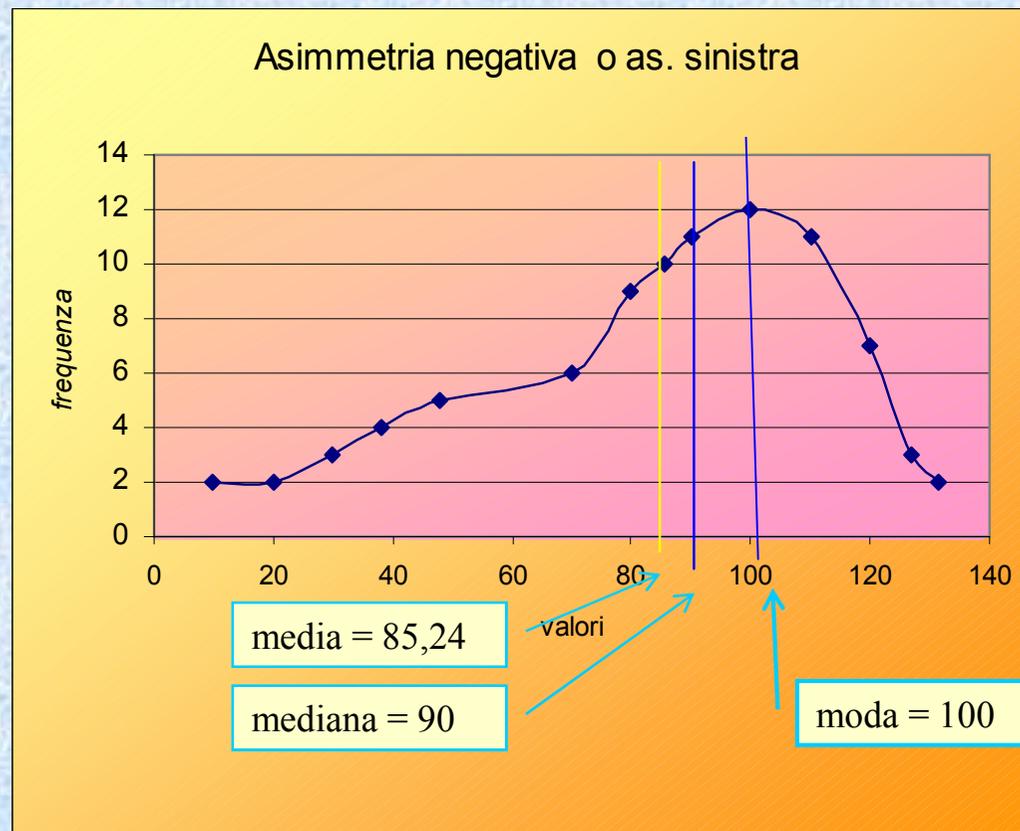


Asimmetria negativa (as. Sinistra)

Si ha un'asimmetria negativa o sinistra quando il ramo sinistro della curva è più lungo di quello destro

In questo caso si ha:

$$\text{media} < \text{mediana} < \text{moda}$$



Curtosi

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

Se la curva è

- più appuntita si dice curva Leptocurtica
- più appiattita si dice curva Platicurtica

Coeff. di curtosi di Pearson

σ^2 = scarto quadratico medio

$0 \leq K < +\infty$

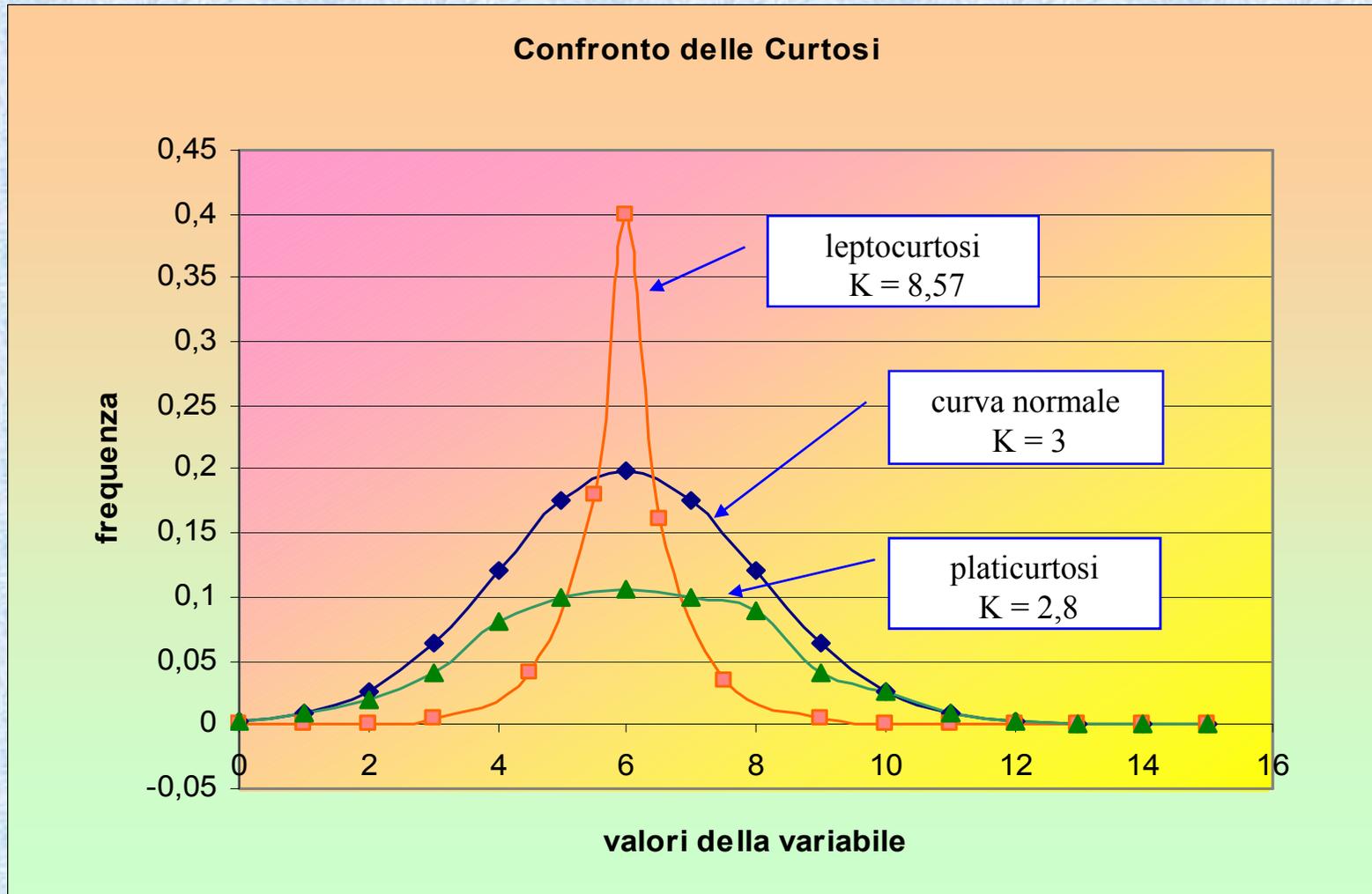
Se $K = 3$ distribuzione normale

se $K > 3$ curva leptocurtica

Se $K < 3$ curva platicurtica.

$$K = \frac{1}{\sigma^4} \left[\frac{\sum_i (x_i - \bar{x})^4 f_i}{\sum_i f_i} \right]$$

Curtosi



Curtosi

Spesso il coeff. di curtosi viene indicato con b_2 che, come visto, nel caso della distribuzione normale è = 3

pertanto, talvolta, la curtosi viene indicata con $(b_2 - 3)$

Allora:

se la distribuzione è normale $(b_2 - 3) = 0$

se la distribuzione è leptocurtica $(b_2 - 3) > 0$

se la distribuzione è platicurtica $(b_2 - 3) < 0$