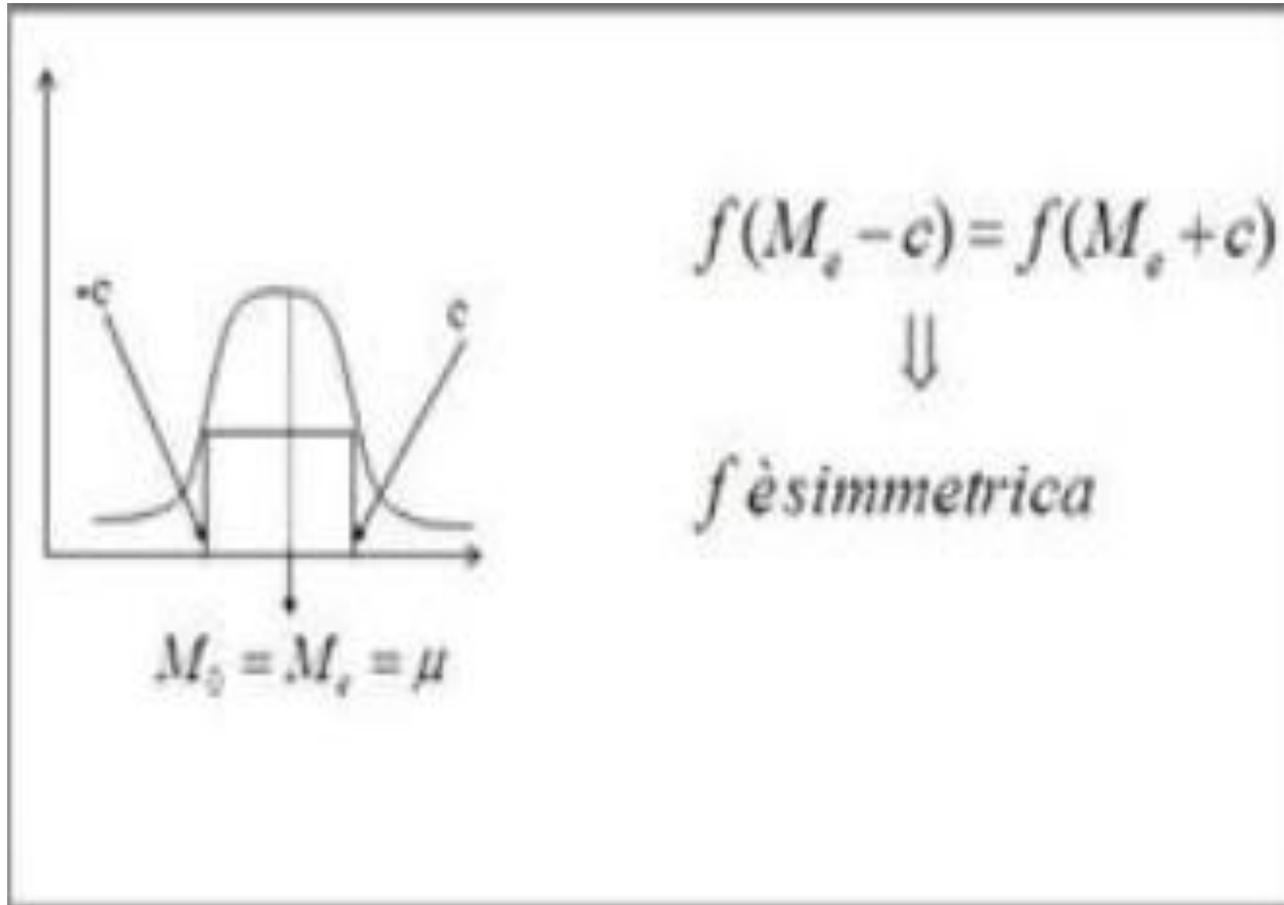


Asimmetria, curva normale e disnormalità

Corso di Statistica I

Prof. Domenico Leogrande

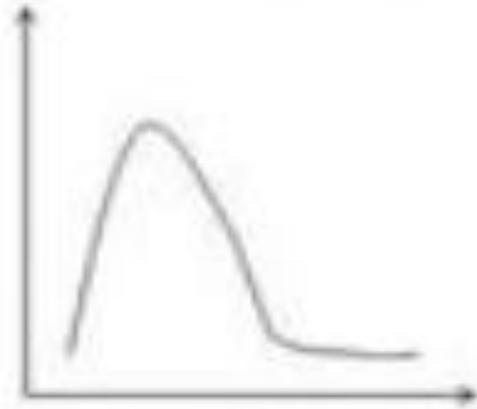
Simmetria



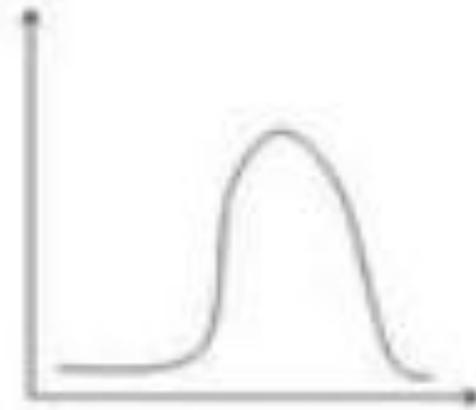
Simmetria intorno al valore mediano: la distribuzione su asse cartesiano presenta una forma che, rispetto al valore mediano, assume una struttura delle frequenze uguale sia nella parte destra che nella sinistra.

Asimmetria

Distribuzioni asimmetriche



Asimmetria positiva o destra



Asimmetria negativa o sinistra

$M_0 < M_e < \mu \Rightarrow$ *asimmetria positiva (destra)*

$\mu < M_e < M_0 \Rightarrow$ *asimmetria negativa (sinistra)*

Indici di asimmetria

$$S_k = \frac{\mu - M_o}{\sigma}$$

$S_k > 0$ Asimmetria positiva

$S_k = 0$ Simmetria

$S_k < 0$ Asimmetria negativa

$$\gamma_1 = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N \sigma^3}$$

$$\gamma_1 = \frac{\sum_{i=1}^N (x_i - \mu)^3 n_i}{N \sigma^3}$$

Curva normale

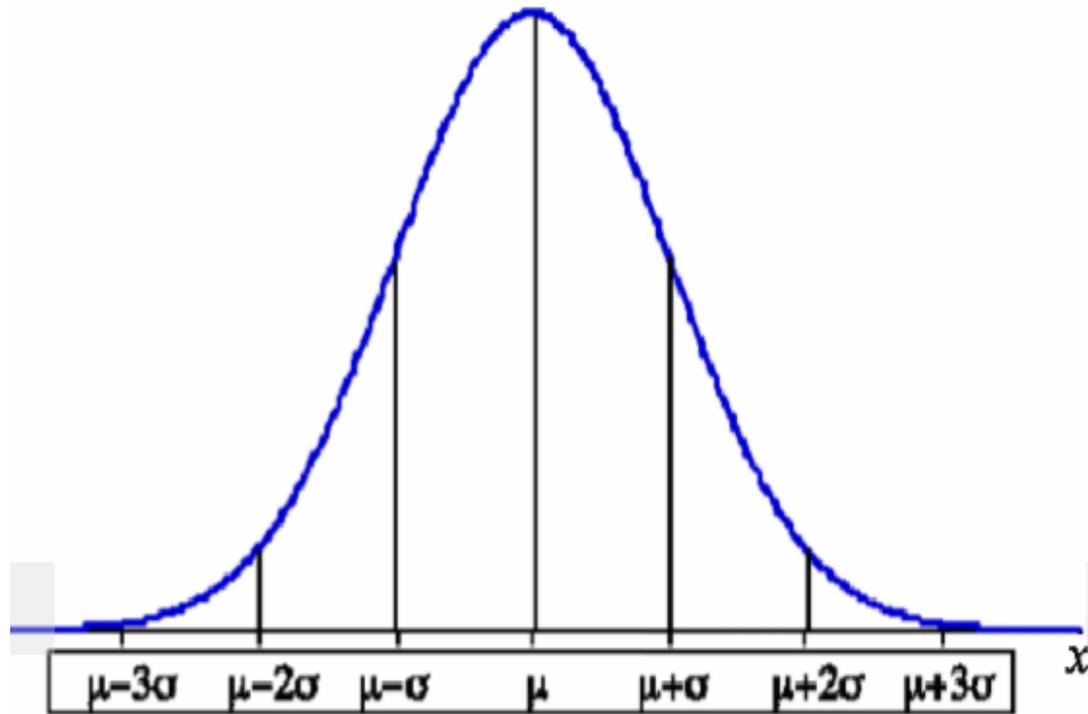
- È la curva continua rappresentativa delle distribuzioni che si incontrano più di frequente in statistica;
- Applicazioni: indagini campionarie, ricerche sperimentali e sociali, campo biomedico.
- La maggior parte delle distribuzioni o variabili stat. quantitative e di tipo continuo tendono a distribuirsi secondo una curva normale, ossia le singole osservazioni di un fenomeno collettivo tendono ad addensarsi intorno al valore medio della osservazione stessa.

Espressione algebrica della curva normale

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

$$\int_{-\infty}^{\infty} \frac{N}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = N$$

Proprietà della distribuzione normale



- È simmetrica rispetto ad μ ;
- è asintotica all'asse delle x da entrambi i lati;
- $\mu = Mo = Me$;
- è crescente per $x < \mu$ e decrescente per $x > \mu$;
- μ è l'ascissa del p.to di max;
- σ è la distanza del p.to di flesso dal p.to di max;
- Possiede due punti di flesso per $\mu - \sigma$ e $\mu + \sigma$;
- L'area sotto la curva è pari a N .

La variabile standardizzata Z

- Standardizzare una v.s.= trasformare la distribuzione originaria in una distribuzione che non risenta dell'effetto scala di misura e dell'effetto media;
- Per finalità pratiche, si riconducono le infinite curve normali caratterizzate da diversi valori di N , μ , σ ad un'unica curva;
- Si considerano le frequenze relative;
- Si considera la variabile standardizzata $z = \frac{x-\mu}{\sigma}$
- F.ne della curva normale standard. $y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$;

Area della curva normale

$$Fr\{a < X < b\} = \int_a^b \frac{N}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = N \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$Fr\{a < X < b\} = Fr\left\{\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

La funzione $\phi(z)$

$$\phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$\phi(z)$ è la funzione di ripartizione della curva normale standardizzata

$$P(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

- z è negativo: $\phi(z) = 0,5 - P(-z)$
- z è positivo: $\phi(z) = 0,5 + P(z)$

Come si calcola la frequenza con Z?

- Definire X e i valori di μ e σ e l'evento di interesse;
- Calcolare il valore di z ;
- Disegnare la curva normale individuando sul grafico l'area di interesse;
- Calcolare il valore della probabilità (area) consultando la tavola.

Esempio (1)

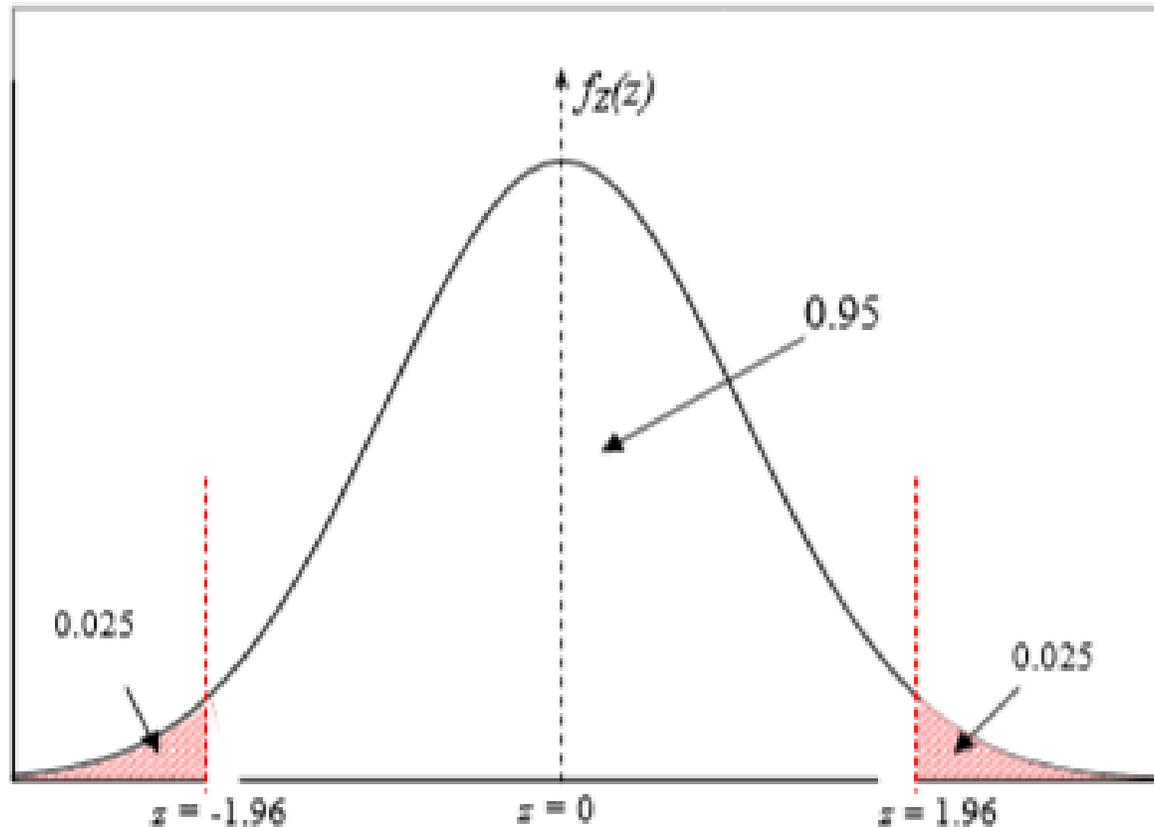
- Calcolare la probabilità $P(Z \leq 3)$:

Dalla tavola B, ottengo che $P(3) = 0,4987$

Per ottenere l'area richiesta, tenendo conto della simmetria della curva, bisogna aggiungere a tale valore 0,5.

Fr $\{z \leq 3\} = 0,5 + 0,4987 = 0,9987$ rappresenta la probabilità che z sia minore o uguale a 3.

Esempio (2)



Calcolare la probabilità
 $P(-1,96 \leq Z \leq 1,96)$:
Dalla tavola B, ottengo che
 $P(1,96) = 0,475$
Per ottenere l'area richiesta,
tenendo conto della
simmetria della curva:
 $Pr\{-1,96 \leq z \leq 1,96\} =$
 $0,475 + 0,475 = 0,95$

Esempio (3)

- Trovare la probabilità $P(Z \leq -0,6)$:

L'area richiesta si trova a sinistra del valore $z=0$; trovo sulla tabella B il valore dell'area compresa tra 0 e 0,6 che risulta pari a 0,2258.

Considerando la simmetria della curva e sapendo che se z è negativo:

$$\phi(z) = 0,5 - P(-z)$$

$$Fr\{z \leq -0,6\} = 0,5 - 0,2258 = 0,2742$$

Esempio (4)

Si assuma che l'altezza delle donne di età maggiore ai 18 anni in Puglia è distribuita secondo la legge normale; la media e lo s.q.m. dell'altezza delle donne sono di 158 cm e 15 cm.

- Calcolare la probabilità che l'altezza di una donna sia compresa tra 140 cm e 170 cm.

$$\text{Fr}\{140 < X < 170\} = \text{Fr}\left\{\frac{140-158}{15} < Z < \frac{170-158}{15}\right\} = \text{Fr}\{-1,2 < Z < 0,8\}$$

$$\phi(-1,2) = 0,5 - P(-z) = 0,5 - 0,3849 = 0,1151$$

$$\phi(0,8) = 0,5 + P(z) = 0,5 + 0,2881 = 0,7881$$

$$\text{Fr}\{-1,2 < Z < 0,8\} = 0,7881 - 0,1151 = 0,673$$

Ciò significa che su un campione di 1000 donne il 67,3% avrà un'altezza compresa tra 140 e 170 cm.

Esempio (5)

- Calcolare la probabilità che l'altezza di una donna sia maggiore di 185cm

$$\text{Fr}\{X>185\} = \left\{Z > \frac{185-158}{15}\right\} = \text{Fr}\{Z>1,8\}$$

$$\phi(1,8) = 0,5 + P(z) = 0,5 + 0,4641 = 0,9641$$

$$\text{Fr}\{Z>1,8\} = 1-0,9641=0,0359$$

Ciò significa che solo il 3,59% delle donne avrà un'altezza superiore ai 185 cm.

Disnormalità

CURTOSI:

- Ipernormale o leptocurtica: la distribuzione è più alta al centro e nelle code, mentre risulta più bassa ai fianchi;
- Iponormale o platicurtica: la distribuzione è più bassa della curva normale al centro e nelle code, mentre è più spessa nei fianchi.

Curtosi

COEFFICIENTE DI ECCESSO DI CURTOSI DI PEARSON:

$$\gamma_2 = \frac{\sum_{i=1}^N (x_i - \mu)^4}{N\sigma^4} - 3 \qquad \gamma_2 = \frac{\sum_{i=1}^S (x_i - \mu)^4 n_i}{N\sigma^4} - 3$$

$\gamma_2 = 0$ curve normali

$\gamma_2 < 0$ distribuzione iponormale

$\gamma_2 > 0$ distribuzione ipernormale

INDICE DI DISNORMALITA' DI GINI

$$I = \frac{2\sigma^2}{\delta^2} - \pi$$

$I = 0$ curve normali

$I < 0$ distribuzione iponormale

$I > 0$ distribuzione ipernormale