L'EVOLUZIONE DELLE VALUTAZIONI IMMOBILIARI

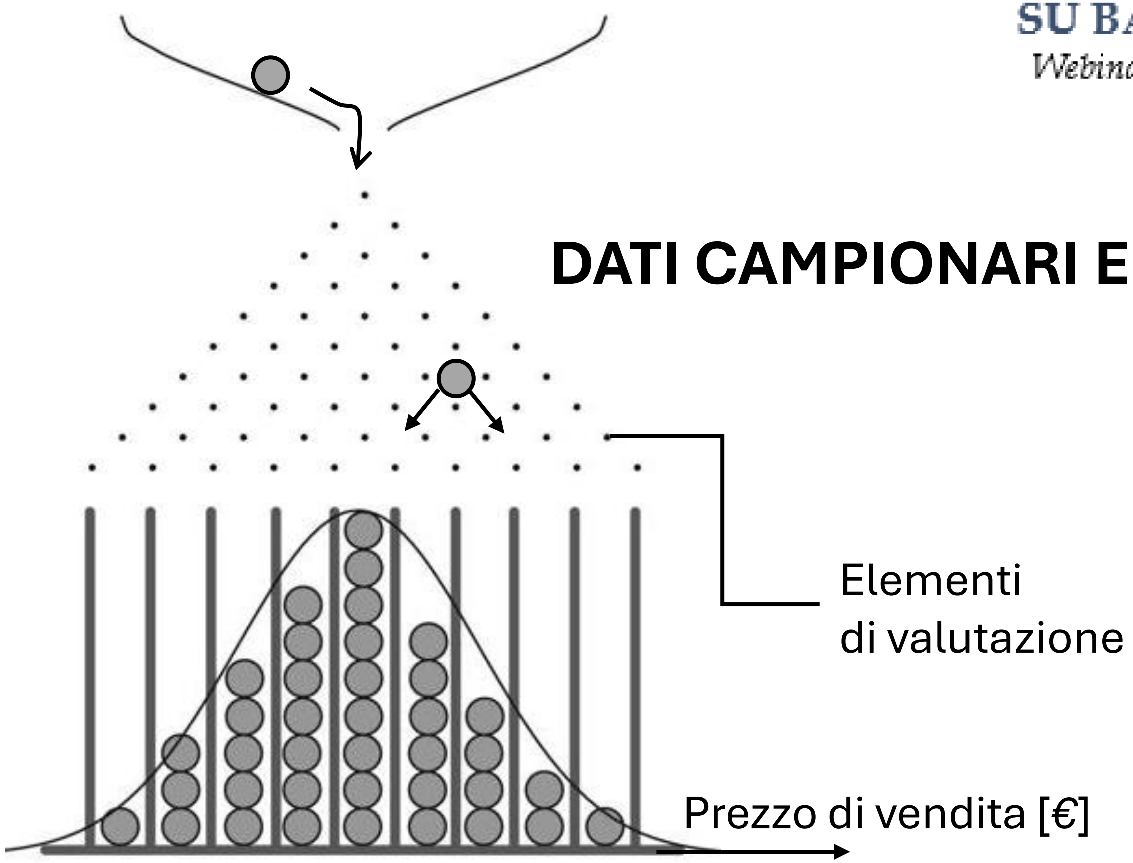
Nuovi paradigmi e modelli interpretativi

LE VALUTAZIONI SU BASE COMPARATIVA

Webinar – 24 ottobre 2024 (15.00/19.00)

DATI CAMPIONARI E DISTRIBUZIONI DI PROBABILITÀ

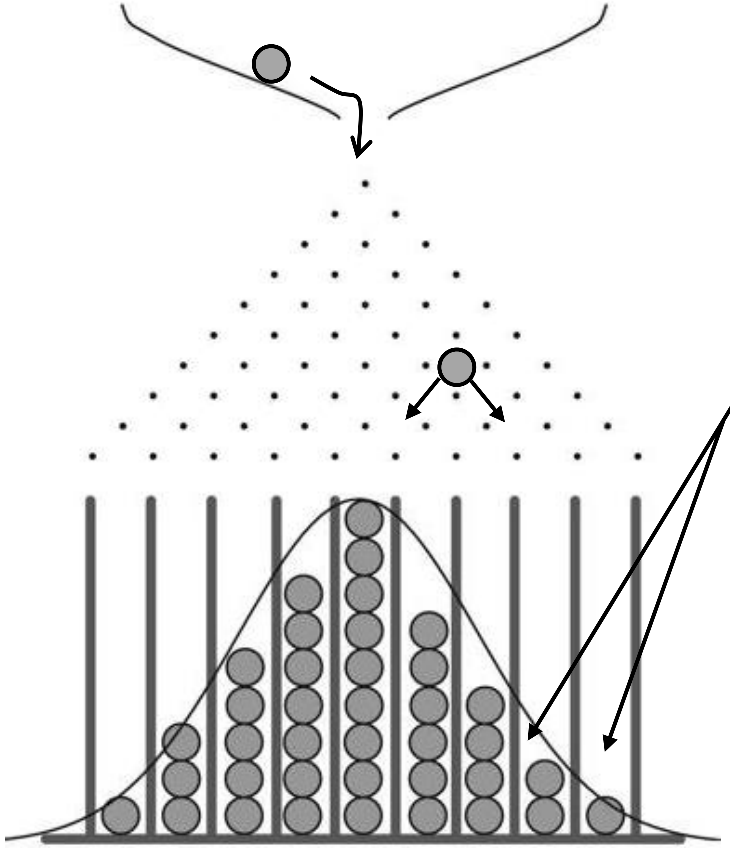
Luca G. Lanza
Università degli Studi di Genova
luca.lanza@unige.it



Fonte: https://www.treccani.it/enciclopedia/tavola-di-galton_(Enciclopedia-della-Matematica)/

TAVOLA di GALTON

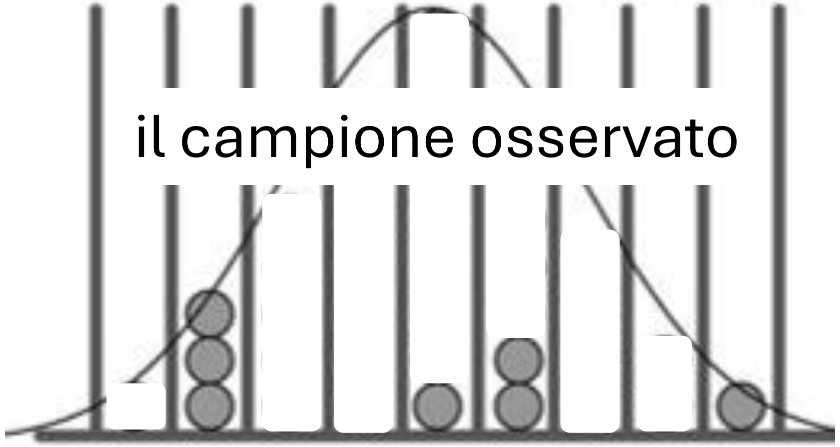
VARIABILE CASUALE: POSIZIONE RAGGIUNTA



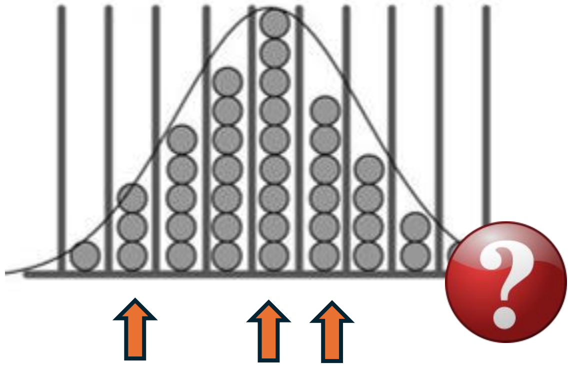
Legge del CASO

Le palline cadono anche dove è meno probabile (ma in numero minore)

Conosciamo la posizione di alcune (poche) palline soltanto



Normalmente non sappiamo quale sia la posizione più probabile ...



né la dispersione delle posizioni raggiunte attorno a quella più probabile

ANALISI CAMPIONARIA

Insieme di tutte le realizzazioni (vendite di unità simili) che conosciamo

Possiamo però studiare il campione osservato (x_1, x_2, ... x_n) e ricavarne alcune informazioni sintetiche.

Se le osservazioni sono esprimibili come numeri reali (ad es. valori immobiliari)

- Media

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

- Dispersione (deviazione standard)

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n-1} - \bar{x}^2 \cdot \frac{n}{n-1}}$$

➢ Mediana

Una volta ordinati i valori osservati: $y_1, y_2, \dots\dots, y_N,$

- per **N dispari** la quantità $m = y_{\frac{N+1}{2}}$, ovvero il termine che occupa il posto centrale della distribuzione $(posizione\ \frac{N+1}{2})$

- per **N pari** la quantità $m = \frac{y_{\frac{N}{2}} + y_{\frac{N}{2}+1}}{2}$, ovvero la media aritmetica dei termini centrali $(posizione\ \frac{N}{2}\ e\ \frac{N}{2} + 1)$

$\begin{cases} \text{50\% valori più alti} \\ \text{50\% valori più bassi} \end{cases}$

➢ Moda

Valore che presenta la probabilità (o frequenza) più elevata

Distribuzione di probabilità

Una distribuzione di probabilità è un modello matematico che collega i valori di una variabile casuale X alla probabilità *P* che tali valori possano essere osservati.

$P = 0$ → Evento impossibile
$P = 1$ → Evento certo (deterministico)

In base alla tipologia di variabile osservata si possono avere:

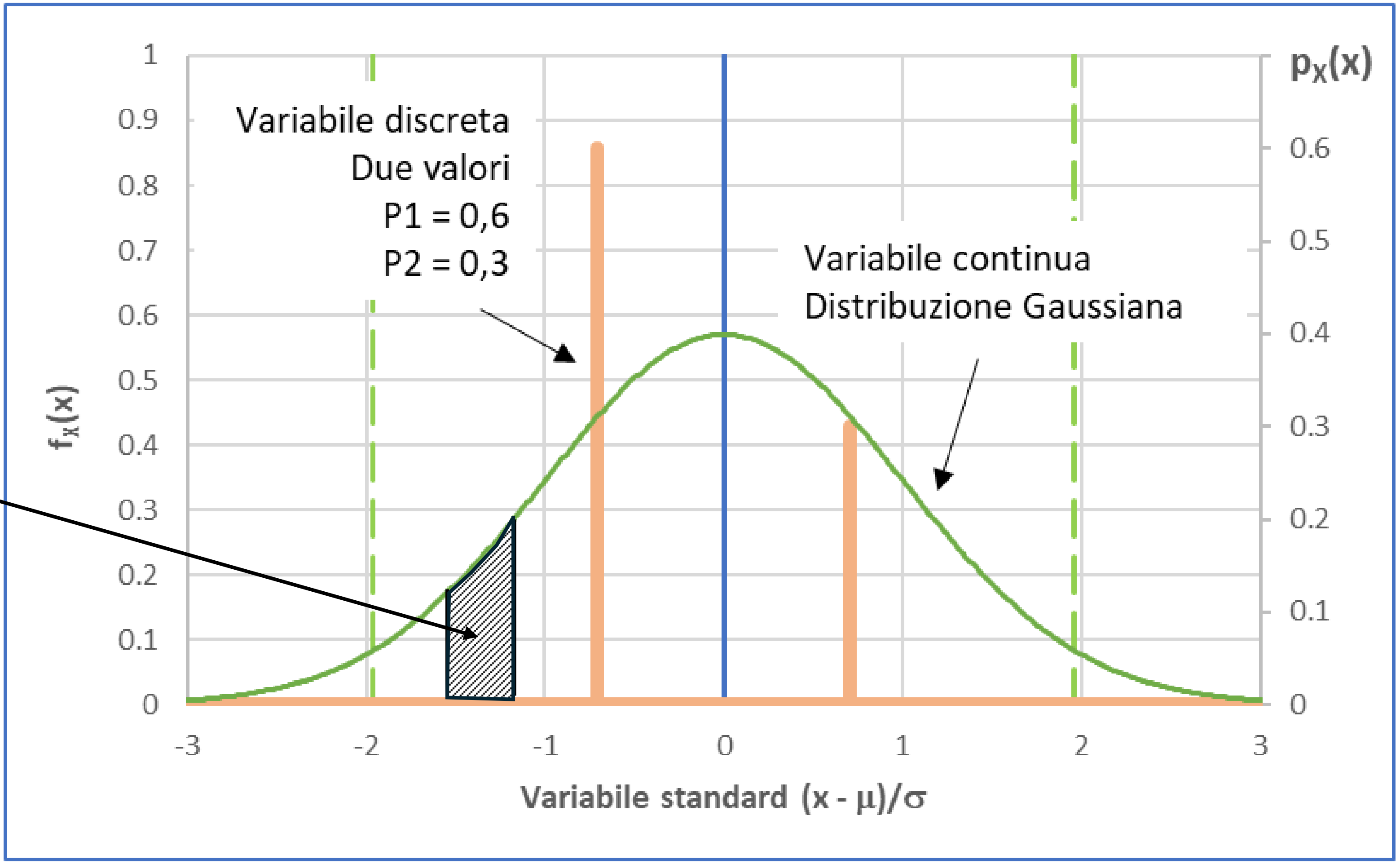
- Distribuzioni **discrete** (la variabile è misurata attraverso valori numerici interi – es. numero di prodotti difettosi in una produzione industriale);

- Distribuzioni **continue** (la variabile è misurata su una scala continua – ad es. il diametro di un disco metallico)

Le distribuzioni di probabilità vengono espresse da una legge matematica detta:

- Funzione di probabilità, indicata con $\boldsymbol{p_X(x)}$, per le variabili discrete
- Funzione di densità di probabilità, indicata con $\boldsymbol{f_X(x)}$, per le variabili continue

L'area sottesa tra due valori indica la probabilità che la variabile casuale assuma un valore compreso tra questi



NB: l'area sottesa complessiva è pari a 1

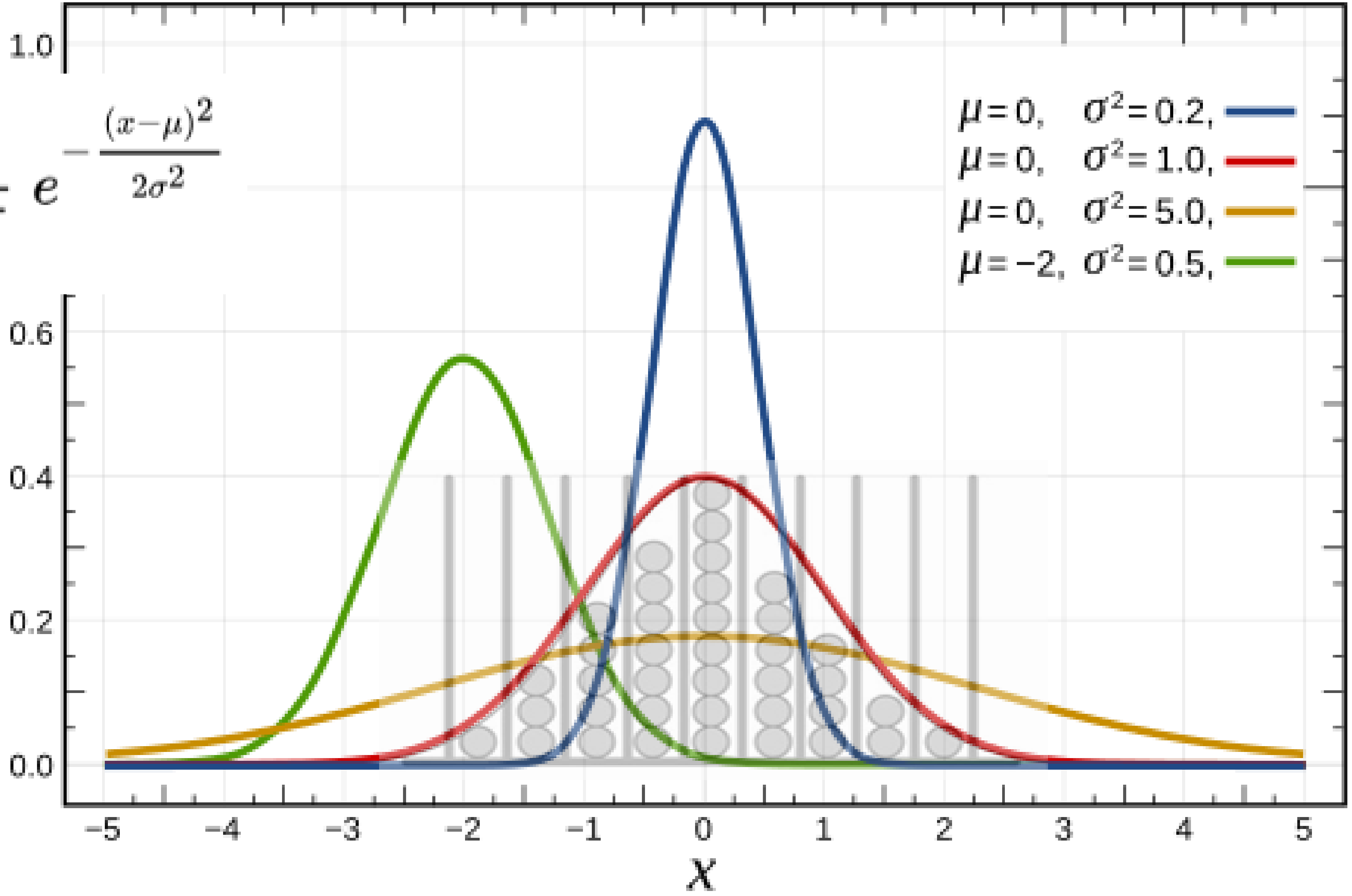
La **distribuzione normale** (o **distribuzione di Gauss** dal nome del matematico tedesco) è una distribuzione di probabilità continua che è spesso usata come prima approssimazione per descrivere variabili casuali che tendono a concentrarsi attorno a un singolo valor medio.

Il grafico della funzione di densità di probabilità associata è simmetrico e ha una forma a campana, nota come "curva a campana", "curva normale" o "curva gaussiana".

La distribuzione normale dipende da due parametri, la media e la varianza (quadrato della deviazione standard), ed è indicata tradizionalmente con:

$$N(\mu; \sigma^2)$$

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



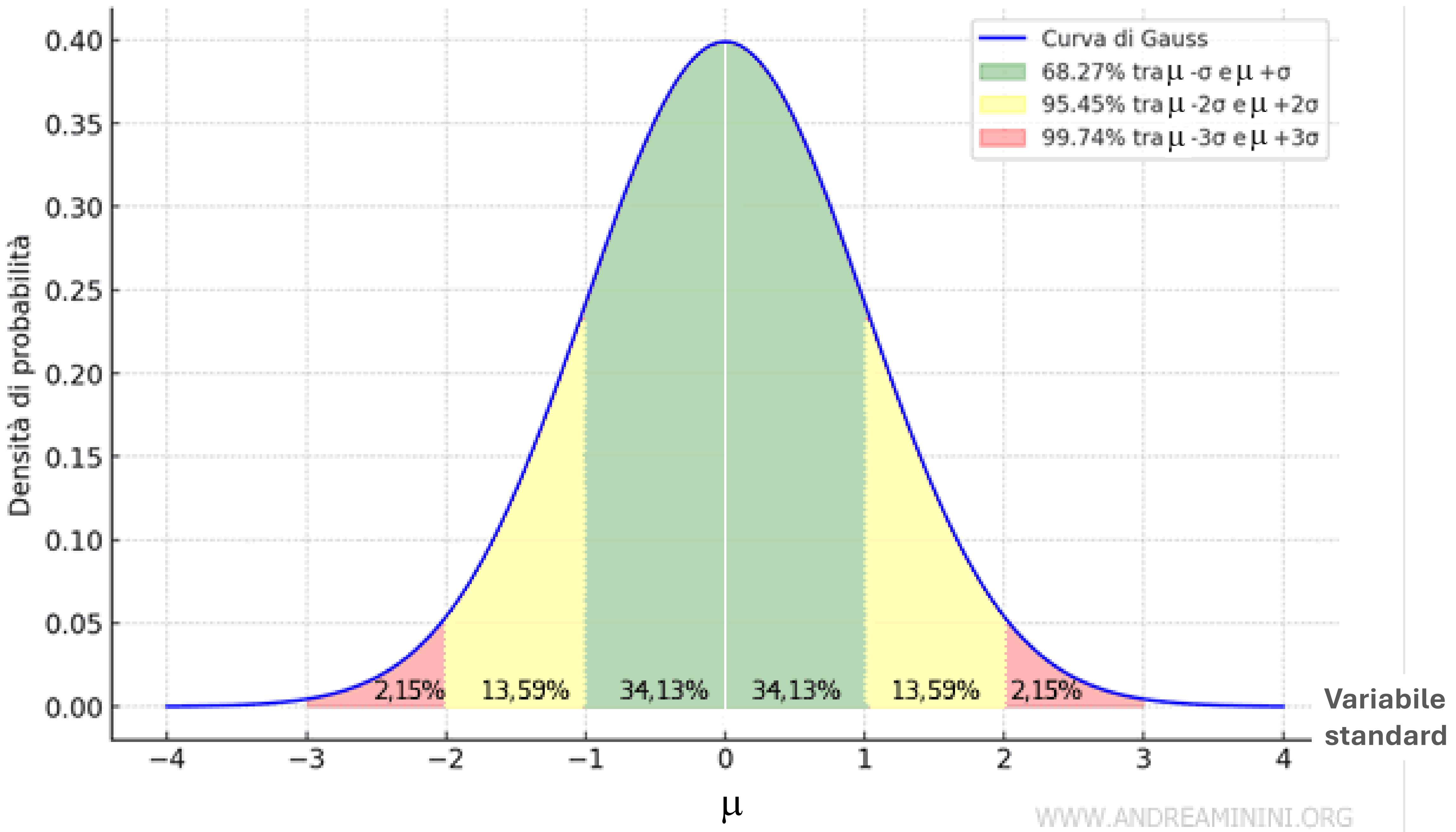
assumendo:

$$Z = \frac{X - \mu}{\sigma}$$

distribuzione Normale Standard N(0,1)

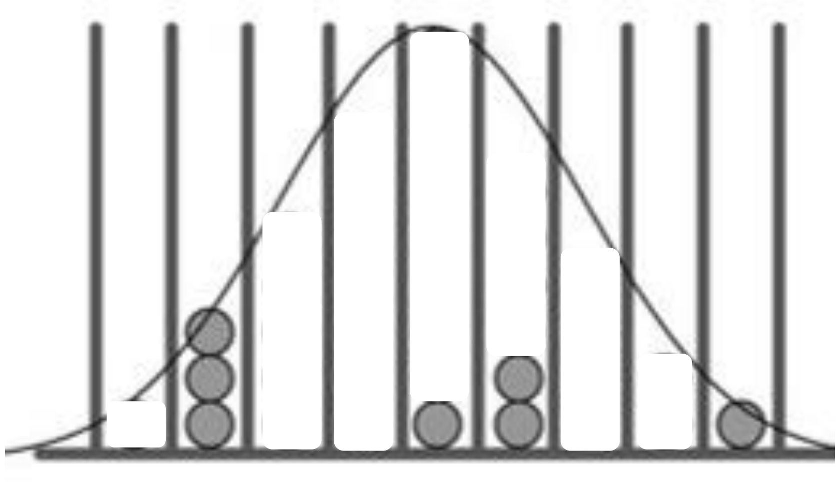
(tabellata e disponibile come funzione predefinita su fogli di calcolo)

Fonte: Inductiveload - self-made, Mathematica, Inkscape, Pubblico dominio, https://commons.wikimedia.org/w/index.php?curid=3817954



Curva di Gauss
68.27% tra μ -σ e μ +σ
95.45% tra μ -2σ e μ +2σ
99.74% tra μ -3σ e μ +3σ
Densità di probabilità
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2,15%
13,59%
34,13%
34,13%
13,59%
2,15%
−4
−3
−2
−1
0
1
2
3
4
Variabile standard
μ
WWW.ANDREAMININI.ORG

Ma normalmente conosciamo la posizione di alcune (poche) palline soltanto ...



... ad esempio un numero limitato di prezzi di vendita di immobili della tipologia di interesse

Come possiamo stimare i parametri della distribuzione di probabilità da cui si può ritenere che quel campione sia stato estratto ?

L'accuratezza nella stima dei parametri dipende dalla lunghezza del campione (quante palline conosciamo)

Intervallo di confidenza (con σ noto)

(Fonte: https://www.unife.it/scienze/informatica/insegnamenti/statistica-applicata/materiale/levine-capitolo-08.pdf)

La stima puntuale è lo specifico valore assunto da una statistica (ad es. la media), calcolata in base ai dati campionari e che viene utilizzata per stimare il vero valore non noto di un parametro di una popolazione

(media del campione $\bar{X}$ → media della popolazione μ)

Possiamo determinare un intervallo costruito attorno allo stimatore puntuale, in modo tale che sia nota e fissata la probabilità che il parametro della popolazione appartenga all'intervallo stesso.
Tale probabilità è detta livello di confidenza ed è in generale indicato con (1−α)% dove α è la probabilità che il parametro si trovi al di fuori dell'intervallo di confidenza
Quindi la confidenza è il grado di fiducia che l'intervallo possa contenere effettivamente il parametro di interesse

Esempio;

Si consideri un processo industriale di riempimento di scatole di cereali e sia assuma che il peso X delle scatole sia X~N(μ;152) – con σ noto.

Dato un campione casuale di n=25 scatole con peso medio 362.3 grammi si vuole costruire un intervallo di confidenza al 95% per μ. Per la proprietà della distribuzione normale e della media campionaria risulta che

$$P\left(-Z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right) = 1-\alpha$$

quindi un intervallo di confidenza all’(1−α)% per μ è dato da

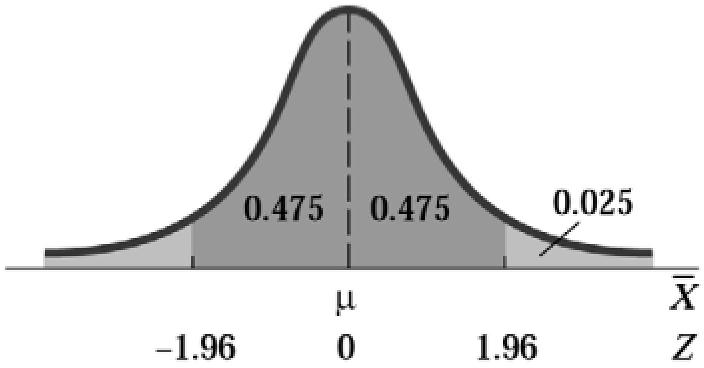
$$\bar{X} - Z_{\alpha/2} \cdot \sigma/\sqrt{n} \leq \mu \leq \bar{X} + Z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

Nel caso specifico si ottiene: 356.42 ≤ μ ≤ 368.18

Nella pratica abbiamo a disposizione un solo campione e siccome non conosciamo la media della popolazione non possiamo stabilire se le conclusioni a cui perveniamo sono corrette o meno.

Tuttavia possiamo affermare di avere una fiducia all'(1−α)% che la media appartenga all'intervallo stimato.

Quindi, l'intervallo di confidenza all'(1−α)% della media con σ noto si ottiene come:



$$\bar{X} - Z_{\alpha/2} \cdot \sigma / \sqrt{n} \leq \mu \leq \bar{X} + Z_{\alpha/2} \cdot \sigma / \sqrt{n}$$

dove $Z_{\alpha/2}$ è il valore a cui corrisponde un'area cumulata pari a (1−α/2) della distribuzione normale standard.

ridurre $Z_{\alpha/2}$

aumentare *n*

Intervallo di confidenza (con σ non noto)

(Fonte: https://www.unife.it/scienze/informatica/insegnamenti/statistica-applicata/materiale/levine-capitolo-08.pdf)

In genere lo scarto quadratico medio della popolazione σ, al pari della media μ, non è noto. Pertanto, per ottenere un intervallo di confidenza per la media della popolazione possiamo basarci sulle sole statistiche campionarie e S.

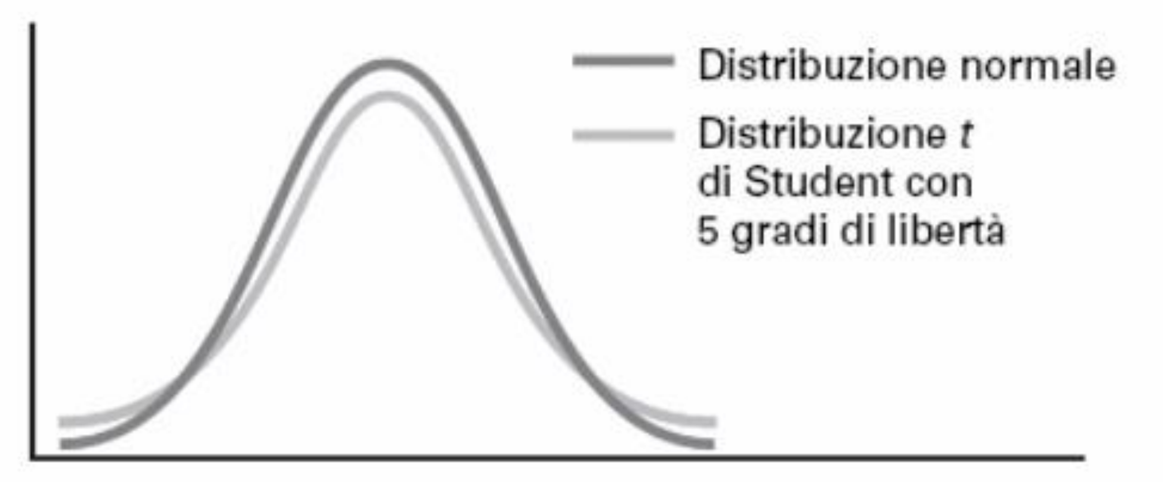
Se la variabile casuale X ha una distribuzione normale allora la statistica:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

ha una distribuzione t di Student con ($n-1$) gradi di libertà (anch'essa tabellata e fornita dai principali fogli di calcolo elettronico).

Se variabile casuale X non ha una distribuzione normale la statistica t ha comunque approssimativamente una distribuzione t di Student in virtù del Teorema del Limite Centrale.

La distribuzione *t* di Student ha una forma molto simile a quella della normale standardizzata. Tuttavia il grafico risulta più appiattito e l'area sottesa sulle code è maggiore di quella della normale a causa del fatto che σ non è noto e viene stimato da S. L'incertezza su σ causa la maggior variabilità di *t*.



All'aumentare dei gradi di libertà, la distribuzione *t* si avvicina progressivamente alla distribuzione normale fino a che le due distribuzioni risultano virtualmente identiche (*n* > *30*).

Il significato dei gradi di libertà è legato al fatto che per calcolare S^2 è necessario calcolare preventivamente $\bar{X}$.
Quindi, dato il valore di $\bar{X}$, solo $n-1$ osservazioni campionarie sono libere di variare: ci sono quindi $n-1$ gradi di libertà.

L'intervallo di confidenza all'($1-\alpha$)% della media quando σ non è noto è definito come:

$$\bar{X} - t_{n-1;\alpha/2} \cdot S/\sqrt{n} \leq \mu \leq \bar{X} + t_{n-1;\alpha/2} \cdot S/\sqrt{n}$$

dove $t_{n-1;\alpha/2}$ è il valore critico a cui corrisponde un'area cumulata pari a ($1-\alpha/2$) della distribuzione I di Student con ($n-1$) gradi di libertà.

In conclusione …

- ✓ Il valore immobiliare può essere interpretato come una variabile casuale derivante dall’interazione di numerosi fattori (noti e meno noti), oggettivi e soggettivi, che ne possono rendere difficoltosa la determinazione.
- ✓ La numerosità e imprevedibilità dei fattori di influenza fanno sì che tale variabile sia descritta da una distribuzione normale (o Gaussiana), i cui parametri tuttavia non sono noti.
- ✓ I parametri della distribuzione si possono stimare da un campione osservato (ad es. i prezzi noti di vendita di alcuni immobili della stessa categoria), ma l’affidabilità della stima dipende dalla numerosità del campione (spesso limitato).
- ✓ Si può calcolare l’affidabilità della stima in base alla numerosità del campione (*n*) e alle statistiche del campione osservato ($\bar{X}, S$).

Il saggio non ha certezze, ha solo ipotesi più o meno probabili. (**Luciano De Crescenzo**)