DISTRIBUZIONI DI PROBABILITA'

Nell'associare ai risultati di un esperimento un valore numerico si costruisce una variabile casuale (o aleatoria, o stocastica).
Ogni variabile casuale ha una corrispondente distribuzione di probabilità.

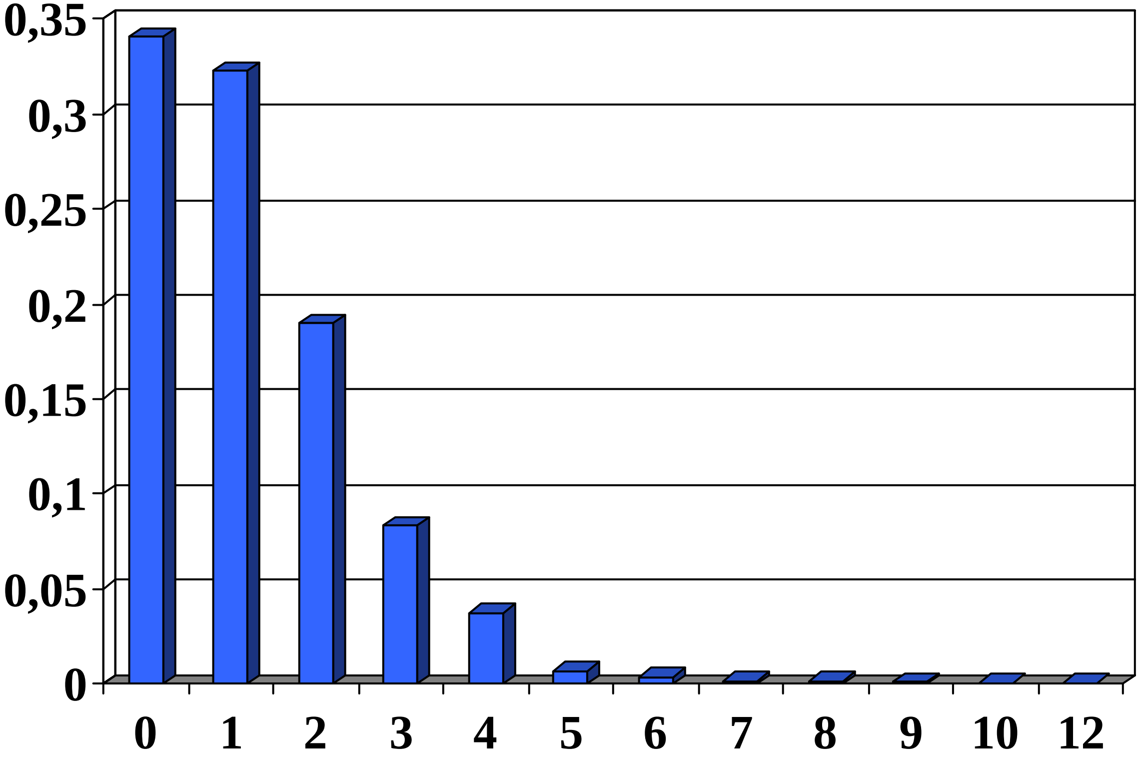
In caso di variabile discreta la distribuzione di probabilità specifica tutti i possibili risultati della variabile insieme alla probabilità che ciascuno di essi si verifichi.

In caso di variabile continua la distribuzione di probabilità consente di determinare la probabilità associata ad intervalli di valori.

Si è analizzato il numero dei farmaci consumati dalle donne gravide ricoverate in un ospedale :
evento A = numero di farmaci
P(A) = probabilità di assumere quel numero di farmaci

N. farmaci	Frequenza	Probabilità P(A)
0	1425	.3405
1	1351	.3228
2	793	.1895
3	348	.0832
4	156	.0373
5	58	.0067
6	28	.0036
7	15	.0014
8	6	.0014
9	3	.0007
10	1	.0002
12	1	.0002
Totale	4185	1

I risultati di una distribuzione di probabilità possono essere riassunti oltre che in tabella anche in grafico.



Si osservi che la probabilità che una donna assuma per esempio 5 farmaci è data dalla frequenza relativa del verificarsi di quel risultato dopo aver eseguito un grande numero di prove (nell'esempio 4185).

→Se X è la variabile casuale che rappresenta tutti i risultati di un possibile esperimento

→ed x è un valore che la variabile X può assumere

→la P(X=x) si definisce attraverso la funzione di probabilità f(x).

Si osservi che è possibile associare ad ogni valore x della variabile X

la $P(X=x)=f(x)\geq 0$

e che la $\Sigma f(x)=1$.

Se X è una variabile casuale con funzione di distribuzione di probabilità f(x) la funzione F(X)=Σf(x) è detta funzione di distribuzione di probabilità cumulata.
La funzione di distribuzione di probabilità cumulativa si ottiene sommando in successione le probabilità $P(X=x_1)+P(X=x_2)+\ldots+P(X=x_n)$

Se X è una variabile continua esiste una funzione f(x) detta funzione di densità di probabilità tale che: $f(x) \geq 0$

$$\int_{-\infty}^{+\infty} f(x)d(x) = 1$$

Per ogni x_1 e x_2 tali che $-\infty \leq x_1 < x_2 \leq +\infty$ allora.

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)d(x)$$

DISTRIBUZIONE BINOMIALE

La distribuzione binomiale è derivata da un processo o esperimento che può assumere solo due risultati mutuamente esclusivi, detto prova di Bernoulli. Una sequenza di prove di Bernoulli forma un processo di Bernoulli sotto le seguenti condizioni:

- **ogni prova può assumere uno dei <u>due</u> possibili <u>risultati, mutuamente esclusivi</u>, uno dei quali è definito arbitrariamente successo e l'altro fallimento;**

- **la <u>probabilità di un successo</u> è denotata con <u>p</u> e <u>rimane costante in ogni prova</u>; la probabilità di fallimento è denotata come q=1-p;**

- **le <u>prove sono indipendenti</u>, cioè il risultato di una prova non influenza il risultato della successiva.**

Si consideri l'esperimento nascita i cui possibili risultati sono

Nascita di sesso F; $p=P(F)$ Nascita di sesso M; $q=P(M)=1-p$.

Si consideri la variabile casuale X=numero di nati di sesso F su n nascite.

Si vuole calcolare la probabilità di osservare 5 nascite F su 8 nascite $P(X=5)$.

Si considera pertanto una sequenza generica di nascite:

F e F e F e M e M e F e F e M.

La probabilità che tale sequenza si verifichi è:

$P(F \cap F \cap F \cap M \cap M \cap F \cap F \cap M)$.

Essendo eventi indipendenti si può scrivere:

$P(F)*P(F)*P(F)*P(M)*P(M)*P(F)*P(F)*P(M)= p*p*p*q*q*p*p*q=p^5q^3$

La sequenza che si può osservare è però una delle tante possibili combinazioni di nascite di 5 femmine e 3 maschi: **FFFFMMMF**

FFMFFMFM …

Per determinare tutte le possibili sequenze si ricorre al calcolo combinatorio:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \binom{8}{5} = \frac{8!}{5!(8-5)!} = 56$$

8 in classe 5.

n!=n*(n-1)*…*1 nell'esempio 8!=8*7*6*5*4*3*2*1

Essendo le sequenze mutuamente esclusive la probabilità di avere la 1° o la 2° sequenza o la …. 56° sequenza è data dalla somma delle probabilità delle singole sequenze:

$$\mathbf{56 * p^5 q^3}$$

Generalizzando, indicando con n il numero di prove e con x il numero dei successi, si ha:

f(x)=P(X=x)

$$f(x) = \binom{n}{x} * p^x q^{n-x}$$

Per x=0,1,2,3,….,n

f(x) = 0 altrove

Si dimostra che la distribuzione binomiale, spesso indicata con il simbolo B(n, p), ha:

Media = E(X) = μ = np

Varianza = σ^2 = npq = np(1-p)

DISTRIBUZIONE DI POISSON

La distribuzione di Poisson viene utilizzata

- **per variabili discrete**

- **per eventi distribuiti casualmente nel tempo e nello spazio**

E’ una distribuzione molto usata in campo biologico e medico:

- conta dei batteri sulle piastre

- analisi di elementi in campioni di acqua

- conta dei globuli su un vetrino

- numero di interventi di urgenza in un mese presso un pronto soccorso

Affinché un processo possa essere descritto con una distribuzione di Poisson devono verificarsi le seguenti condizioni:

- in un determinato intervallo gli eventi accadono in modo indipendente - il verificarsi di un evento in un intervallo di tempo o di spazio non influenza la probabilità di verificarsi di un secondo evento nello stesso intervallo di tempo o spazio;

- la probabilità di un evento in un intervallo di tempo Δt infinitamente piccolo è direttamente proporzionale alla lunghezza dell'intervallo;

- in una parte infinitamente piccola dell'intervallo la probabilità che più di un evento si verifichi è trascurabile

Se X è la variabile casuale che segue una distribuzione di Poisson la sua funzione di distribuzione di probabilità è

$$f(x) = \frac{e^{-\lambda} * \lambda^{x}}{x!}$$

per x=0,1,2,….., ∞

f(x) = 0 altrove

- "*e*" è la costante 2,7183

- "λ" è il parametro della distribuzione di Poisson e corrisponde alla media.

Una proprietà importante della distribuzione di Poisson è che media e varianza coincidono:

$$\lambda = \mu = \sigma^2$$

E’ possibile approssimare la distribuzione binomiale alla distribuzione di Poisson quando n, il numero delle prove, è molto grande e p, la probabilità di successo, tende a zero:

$n \to \infty$ $p \to 0$ $B(n, p) \approx \text{Poisson}(\lambda)$

$n * p = \lambda = \text{costante}$

DISTRIBUZIONE NORMALE (GAUSS)

Quando si esegue un esperimento e si descrivono i risultati si costruisce spesso un grafico (istogramma) per mostrare l’andamento del fenomeno in esame.

In un istogramma:

- sull’asse delle ascisse (x) poniamo i valori della variabile

- sull’asse delle ordinate (y) poniamo le frequenze con le quali un determinato valore, un intervallo di valori in caso di variabili continue, si presenta.

L’area delle colonnine di un istogramma rappresenta la frequenza con cui i valori x_1 e x_2 che delimitano la base della colonnina si presentano nel nostro esperimento

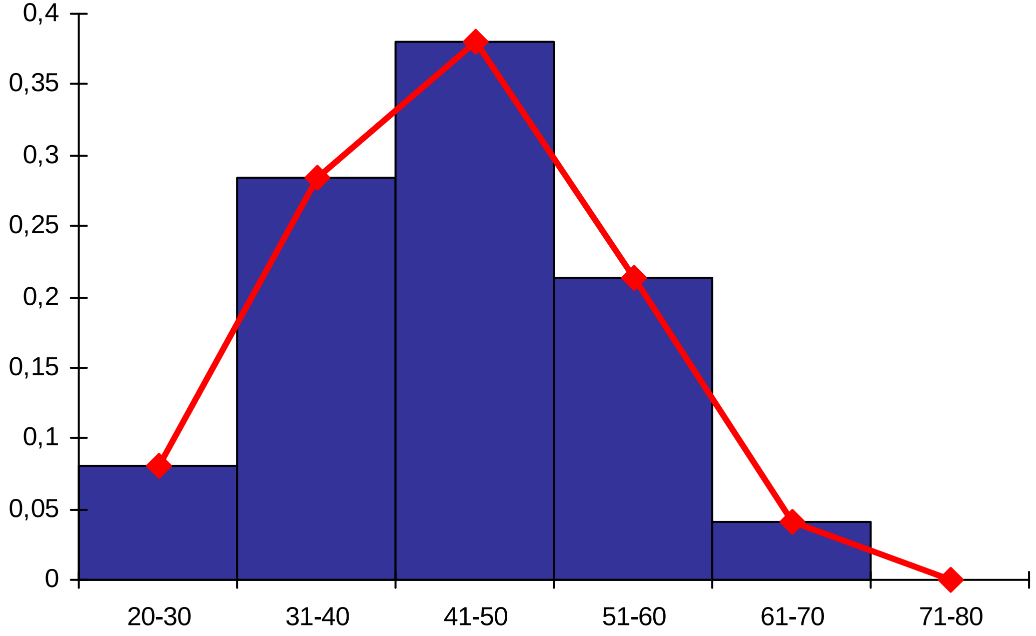
Unendo i punti medi delle classi con una spezzata si costruisce un poligono di frequenza.

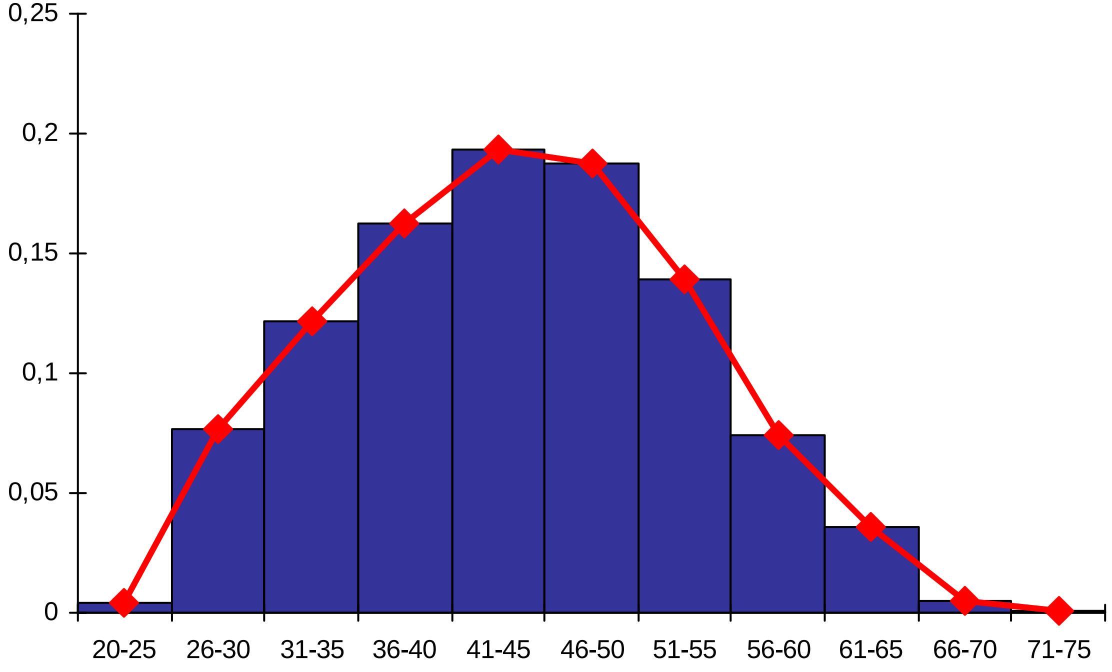
Se scegliamo l’intervallo tra i valori sempre più piccolo e con una quantità di prove molto grandi la forma del poligono di frequenza si avvicinerà sempre più ad una linea continua.

Questa è la rappresentazione grafica della distribuzione di una variabile continua e la funzione che ne è l’espressione matematica è la funzione densità di probabilità

Supponiamo di valutare la distribuzione dei soggetti di una popolazione per età

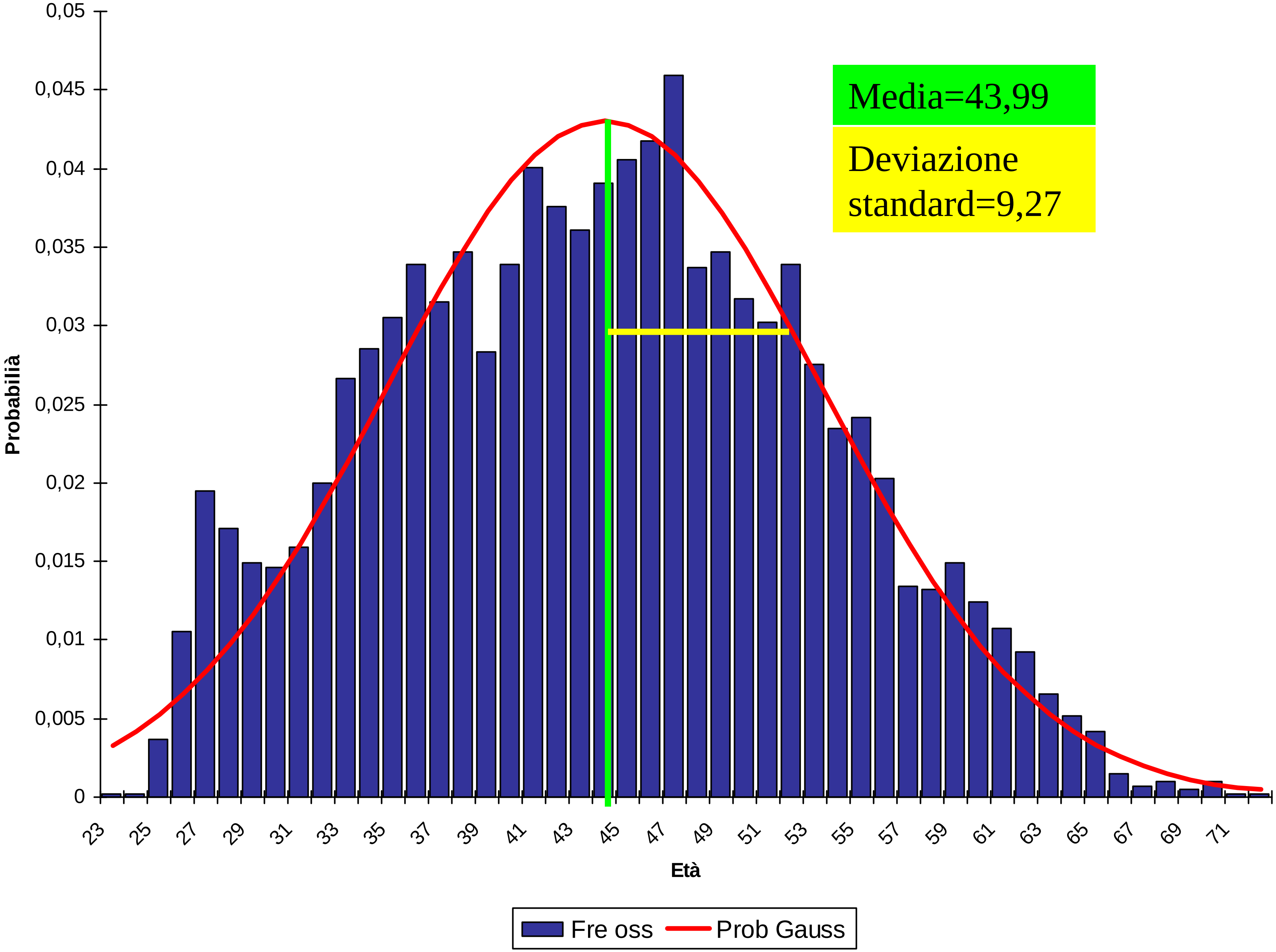
Classe di età	Frequenza assoluta	Frequenza relativa	Frequenza cumulativa	Frequenza cumulativa relativa
20-30	331	8,08%	331	8,08%
31-40	1163	28,38%	1494	36,46%
41-50	1561	38,09%	3055	74,55%
51-60	875	21,35%	3930	95,90%
61-70	166	4,05%	4096	99,95%
71-80	2	0,05%	4098	100,00%
	4098	100,00%		





Possiamo riprodurre la distribuzione di frequenza utilizzando un istogramma. Possiamo inoltre unire i punti medi di ciascuna classe con una linea spezzata per rappresentare il fenomeno con un poligono di frequenza. Se rendiamo l’intervallo di classe progressivamente più piccolo…..

…l'ampiezza può ridursi al punto che il poligono di frequenza possa essere approssimato ad una curva continua



La curva di distribuzione di probabilità di una variabile continua che presenta un andamento “a campana” prende il nome di curva normale o gaussiana.
La sua espressione matematica è

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$-\infty \leq x \leq +\infty$

con

exp = funzione esponenziale

π = 3,14

e i due parametri della distribuzione:

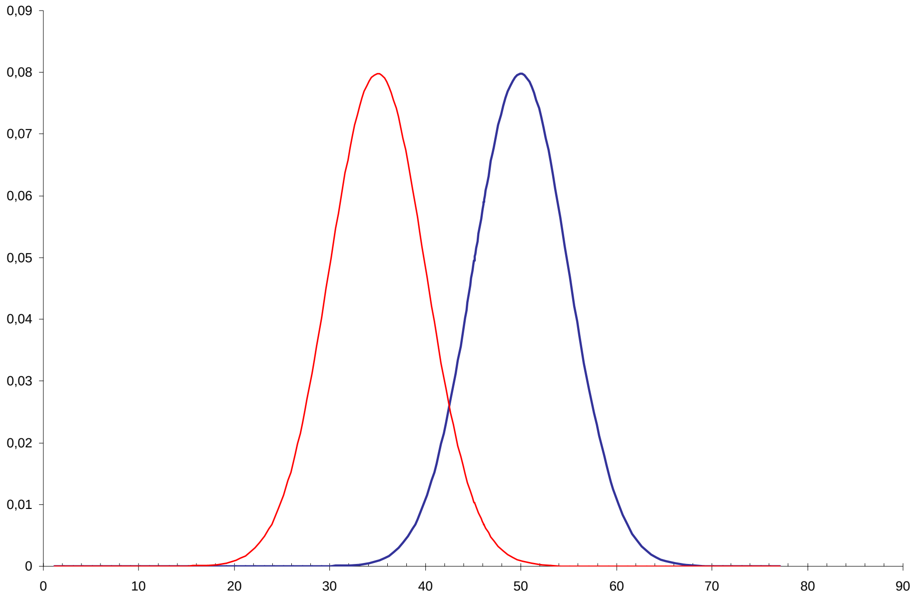
μ = media

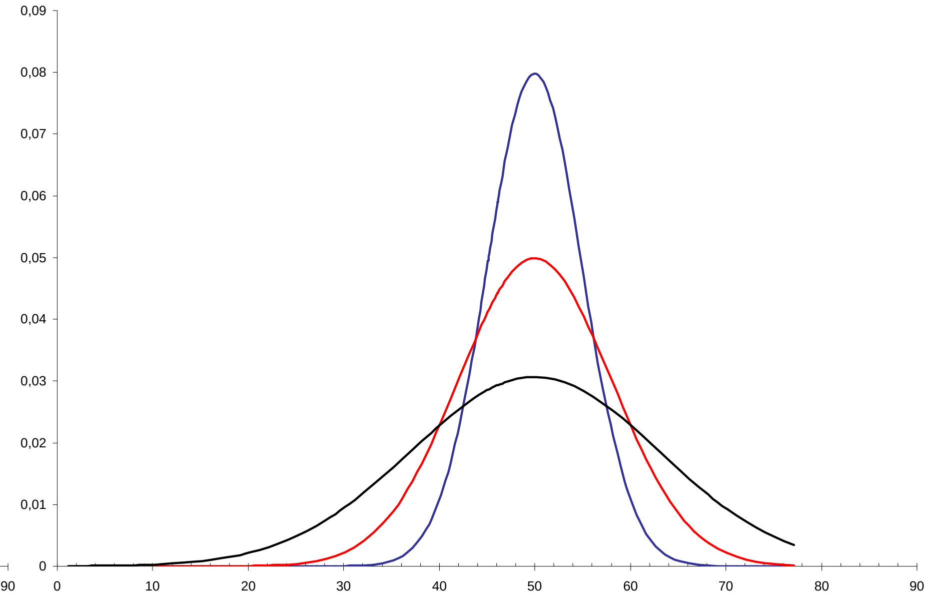
σ = deviazione standard

la distribuzione di Gauss è completamente definita dai valori di μ e σ:

differenti valori di μ spostano la posizione della curva lungo l’asse delle ascisse

differenti valori di σ modificano l’altezza della curva.





La distribuzione di Gauss ha alcune caratteristiche tipiche:

- è simmetrica intorno alla sua media
- la media, la mediana e la moda coincidono
- l’area sotto la curva è uguale ad 1 (100%)
- l’area sotto la curva compresa nell’intervallo
  - μ-σ ed μ+ σ è pari al 68% dell’area totale
  - μ-2 σ e μ+2 σ è pari al 95% del totale
  - μ-3 σ ed μ+3 σ è pari al 99,7% del totale

Esistono degli indici per misurare la normalità della curva di Gauss:

asimmetria:

asimmetria $= 0$ curva normale

asimmetria < 0 coda sinistra pi lunga

asimmetria >0 coda destra più lunga

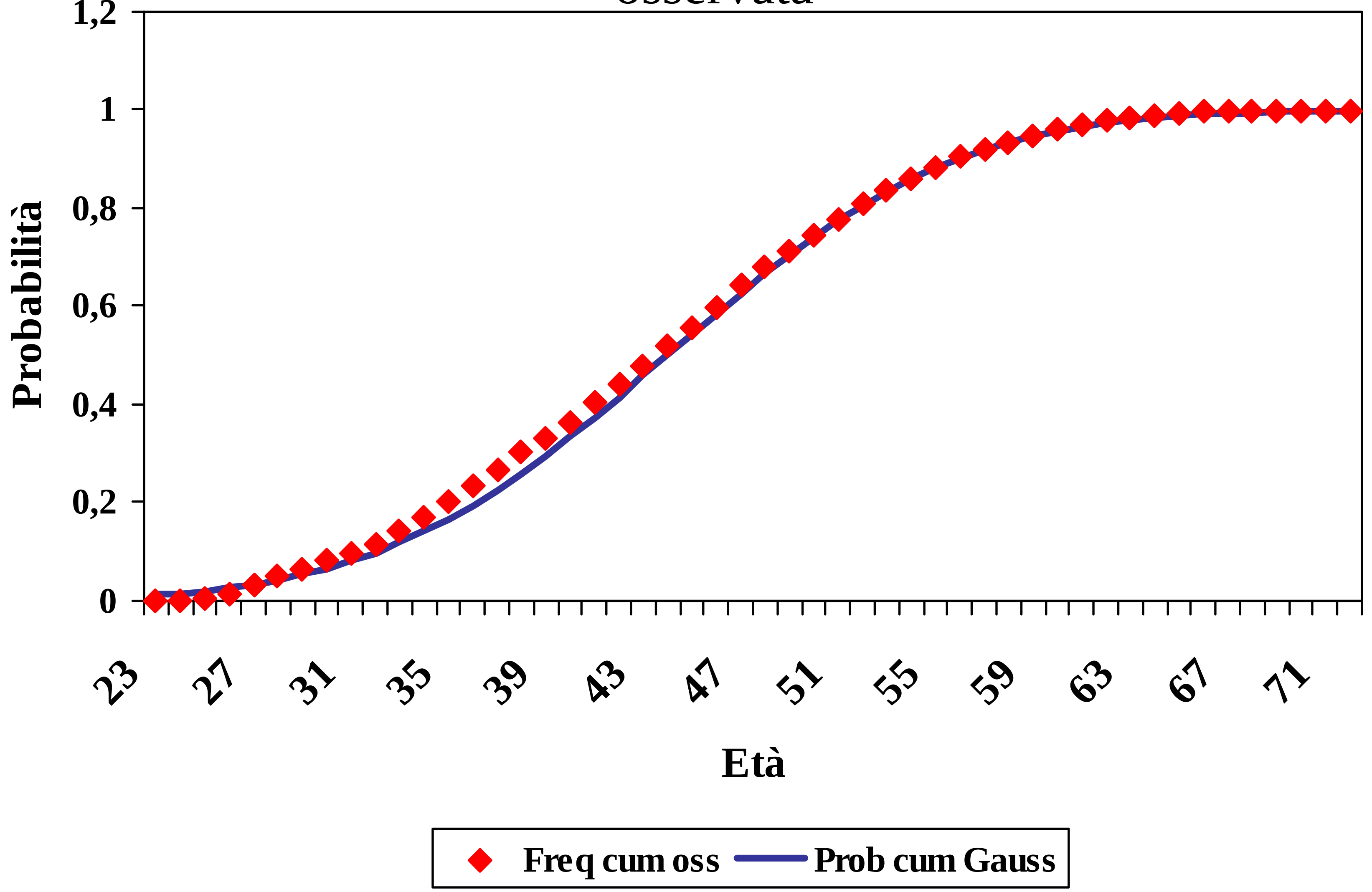
curtosi:

curtosi $= 3$ curva normale

curtosi < 3 code leggere, distribuzione appuntita (ipernormale o leptocurtica)

curtosi > 3 code pesanti, distribuzione piatta (iponormale o platicurtica).

Distribuzione di Gauss cumulativa teorica e distribuzione cumulativa osservata



L'ultima caratteristica enunciata ci dice che esiste una famiglia di distribuzioni di gaussiane ed ogni membro è distinto in base ai valori di μ e σ.

Tra le varie curve di Gauss la più importante è la distribuzione di Gauss standard che ha

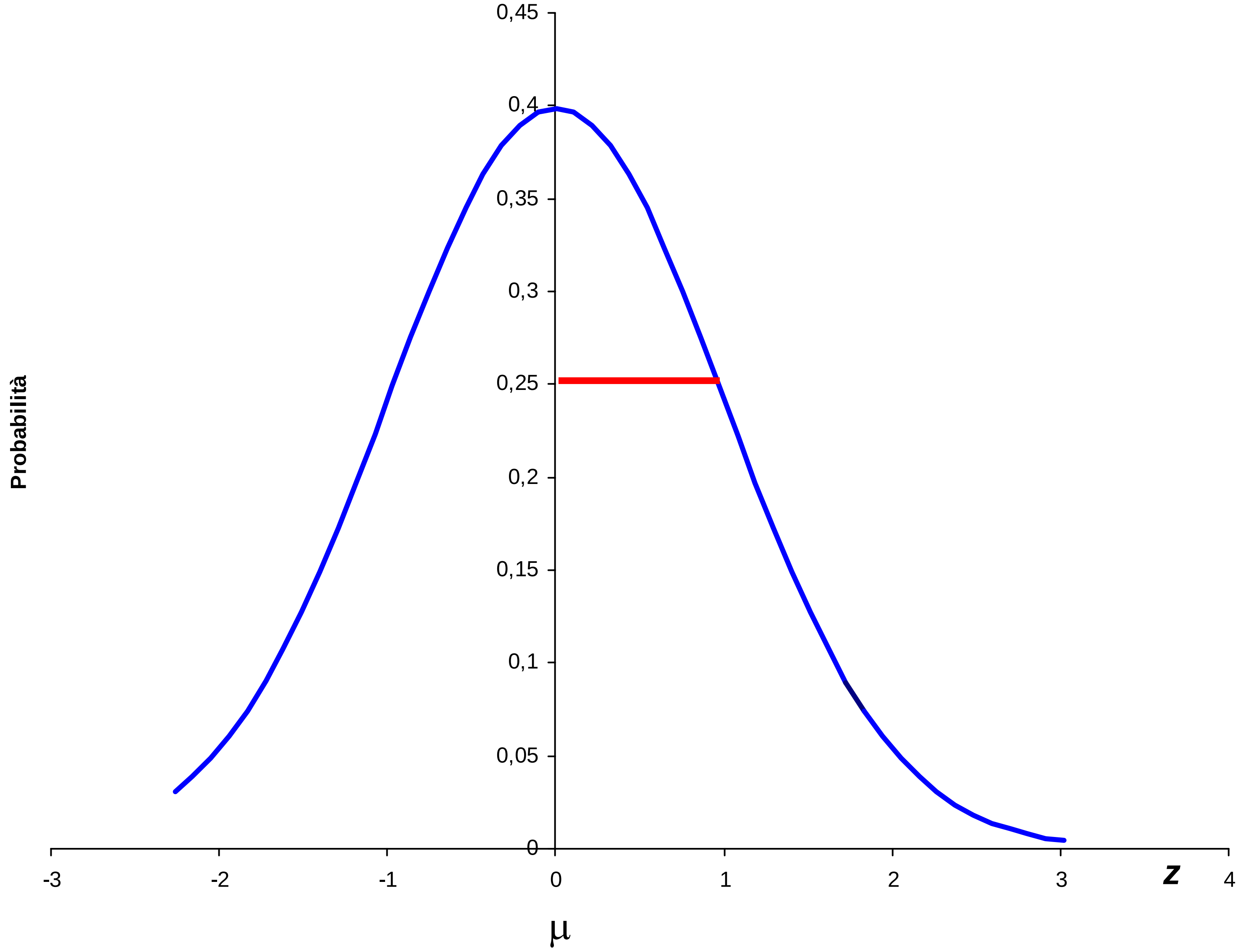
media = 0

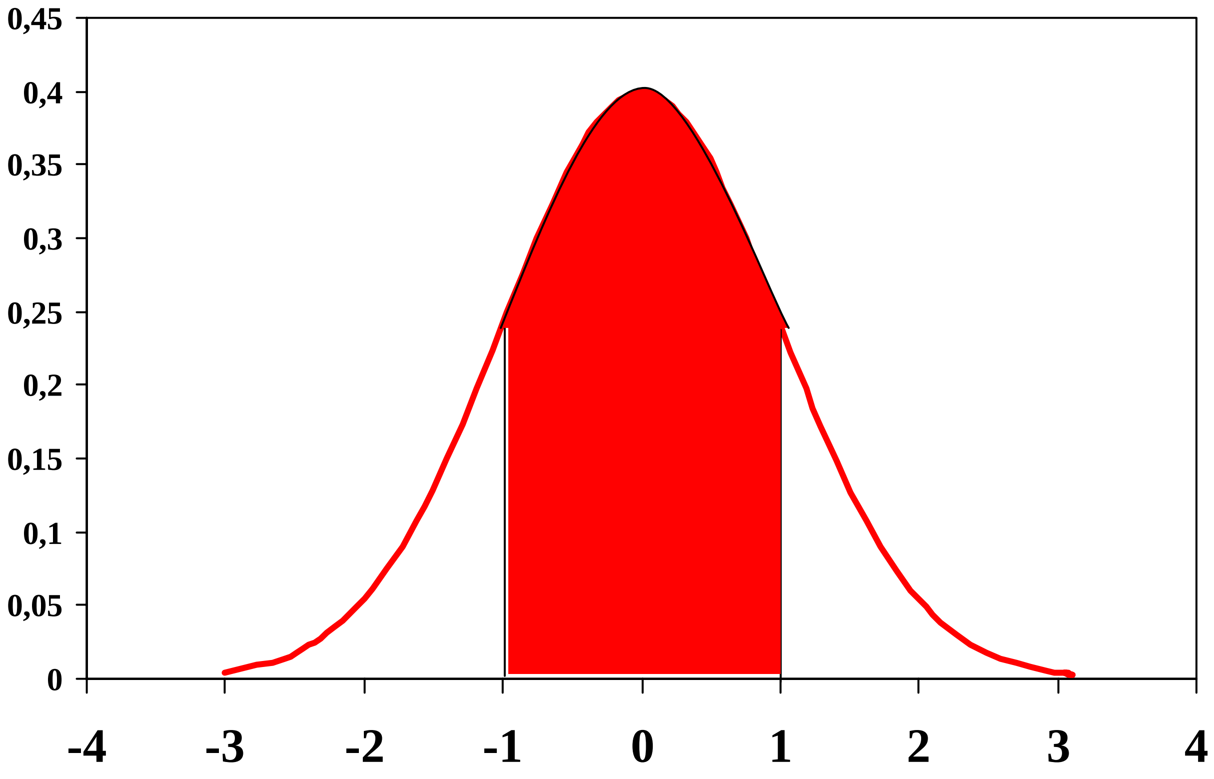
deviazione standard = 1

L'espressione matematica della distribuzione di Gauss Standard è:

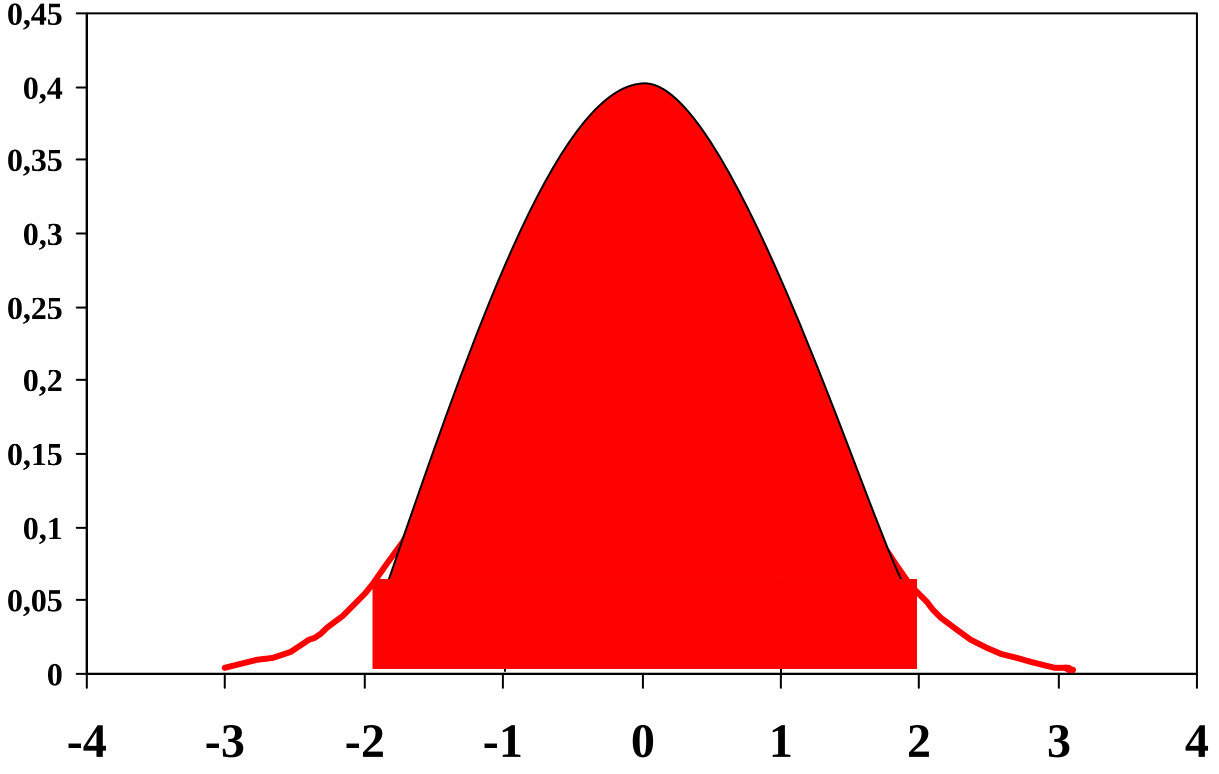
$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{z^2}{2} \right\}$$

Distribuzione di gauss standard: media=0 e deviazione standard=1





Area in rosso = 68%



Area in rosso = 95%

E’ importante sottolineare che ogni variabile X può essere standardizzata mediante la trasformazione

$$z = \frac{x - \mu}{\sigma}$$

Indicando la distribuzione di Gauss con il simbolo $N(\mu, \sigma)$ è possibile fare le seguenti approssimazioni:

$B\ (n, p) \approx N(\mu, \sigma)$ se $\quad n \to \infty \quad p = 0{,}5$

$P\ (\lambda) \approx N(\mu, \sigma)$ se $\quad n \to \infty$