

CDL in INFERMIERISTICA & OSTETRICIA

Statistica Medica

gbarbati@units.it

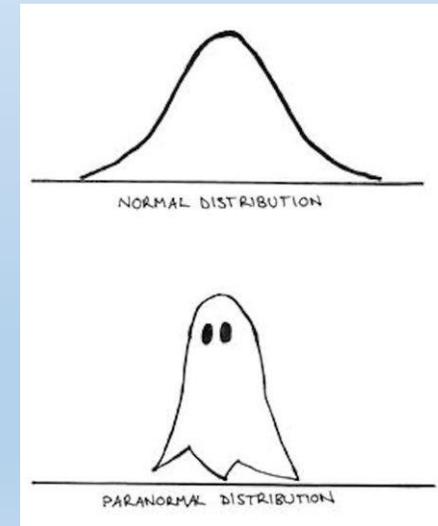
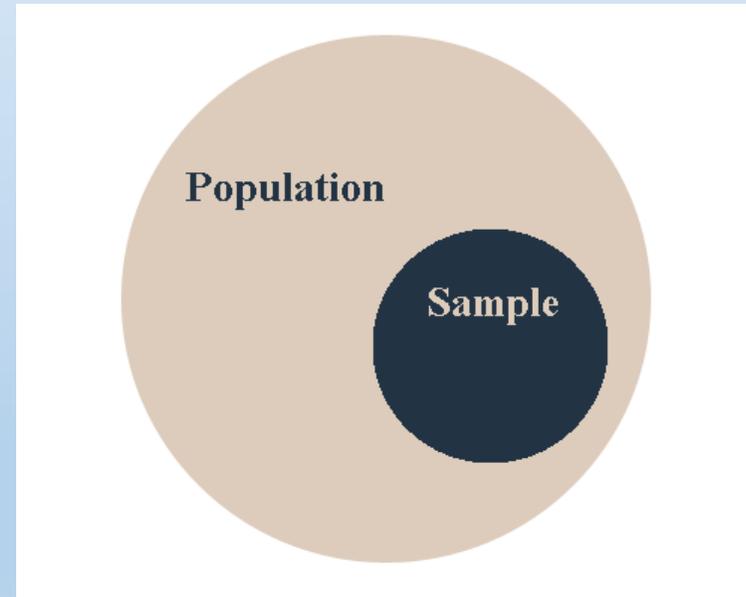
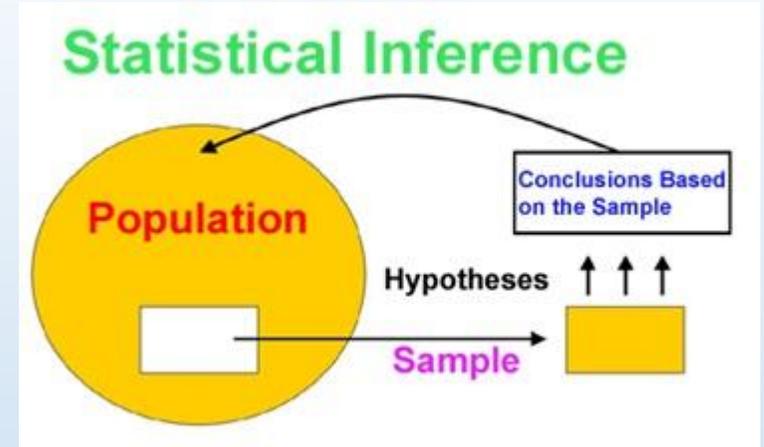
A.A. 2024-25



UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di
Scienze Mediche Chirurgiche e della Salute

Sommario:

- Introduzione all'inferenza
- Definizioni di probabilità
- Cenni di calcolo delle probabilità
- Cenni di calcolo combinatorio
- Variabili Aleatorie & Distribuzioni di Probabilità





Un *trial* clinico ha mostrato che 50 pazienti che ricevono il farmaco A per una malattia guariscono più velocemente (in media) di 50 pazienti che ricevono il farmaco B.

(1) E' giusto affermare, in generale, che A è migliore di B per curare questa malattia?

(2) Il medico dovrà usare in futuro A piuttosto che B per trattare al meglio i suoi pazienti?

(1) Domanda di tipo "**inferenziale**": quali conclusioni possono essere tratte dal **campione** rispetto alla **popolazione** da cui è stato estratto?

(2) Domanda è di tipo "**decisionale**": qual è la scelta più razionale per il futuro trattamento, tenendo conto dell'informazione offerta dal trial?



Le risposte ad entrambe le domande, e a tutte le domande riguardanti studi ***campionari***, sono caratterizzate da un certo livello di ***incertezza***.



L'indicazione dal campione è che il farmaco A sia migliore di B, ma...:

- possiamo essere **certi** che i pazienti che hanno ricevuto B non fossero più gravemente malati di quelli che hanno ricevuto A?
- e che questa **variabilità** tra i due gruppi non fosse il motivo per le loro differenti risposte?

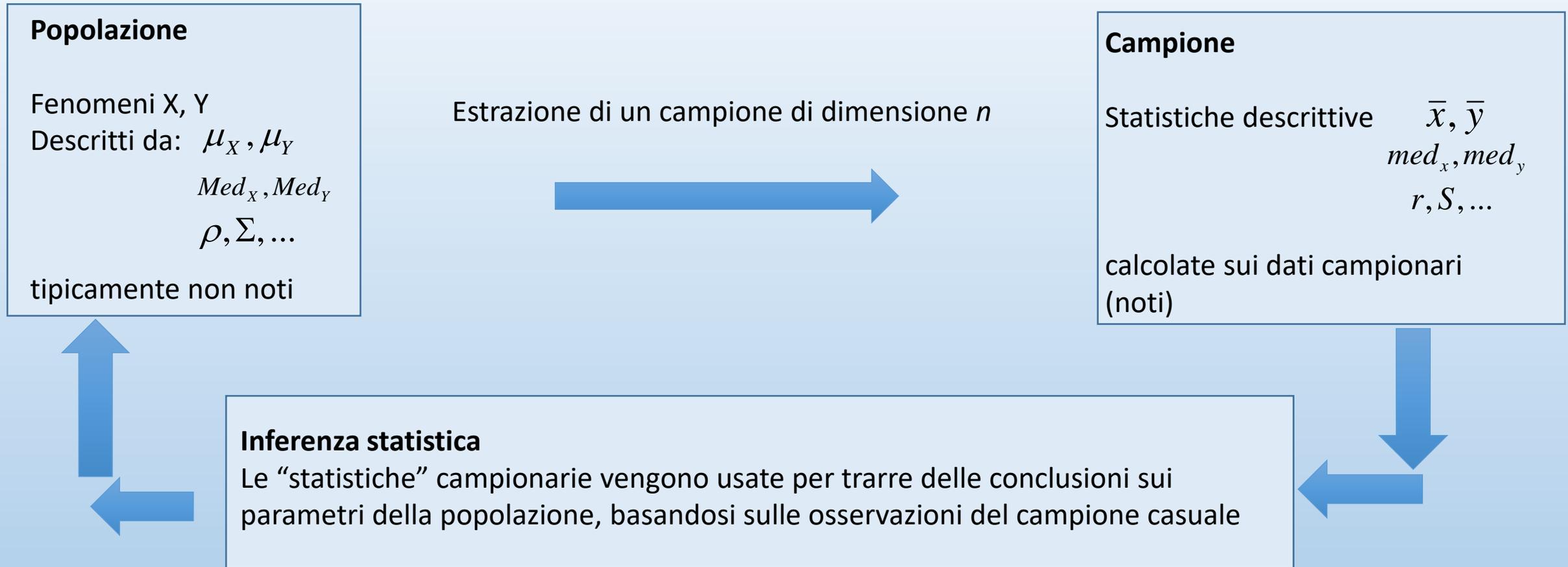
Le risposte devono quindi essere formulate in termini di ***'incertezza'***:

incertezza **bassa** -> conclusione campionaria affidabile, decisione «sicura»

incertezza **alta** -> l'esperimento sarà considerato non conclusivo e non sarà di aiuto nella decisione.

Per avere una *'misura dell'incertezza'* di un risultato campionario, lo strumento matematico appropriato è la ***teoria della probabilità***.

Dalla Statistica Descrittiva alla Statistica Inferenziale*



Quindi:

- stime/conclusioni soggette a **incertezza** (basate su un **campione casuale** della popolazione)
- **teoria della probabilità** è lo strumento matematico per effettuare un'inferenza corretta

* impostazione «frequentista»

Definizione di probabilità

La probabilità rappresenta in termini matematici i fenomeni caratterizzati da comportamenti **variabili**

Se una moneta viene lanciata tante volte e si guarda l'esito di ogni lancio:



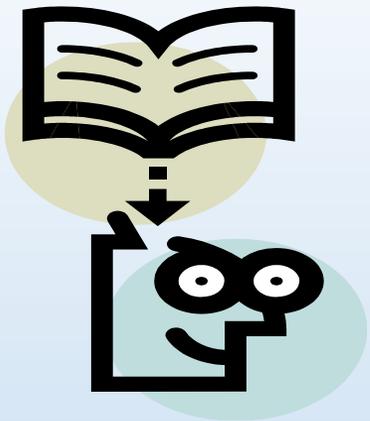
TTCTCCTCTTTCTCCTCCCCTTC....

T=testa; C=croce

Tale sequenza è definita una **sequenza casuale** o una **sequenza random**; ogni posto nella sequenza rappresenta una **prova** o **estrazione (trial)** ed ogni risultato del lancio viene definito come un **evento**.

Una sequenza random è caratterizzata dal fatto che non è possibile predire («*indovinare*») l'evento successivo sulla base dei precedenti.

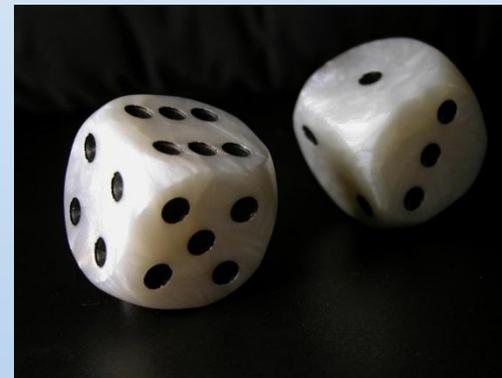
La probabilità di avere 'croce' in un certo passo è la stessa che ad ogni altro, e non è influenzata da ciò che è successo prima



La definizione “classica” di probabilità è:

La probabilità di un evento è data dal rapporto tra il numero dei casi *favorevoli* all’evento ed il numero dei casi *possibili*, purchè questi ultimi siano tutti *ugualmente probabili*.

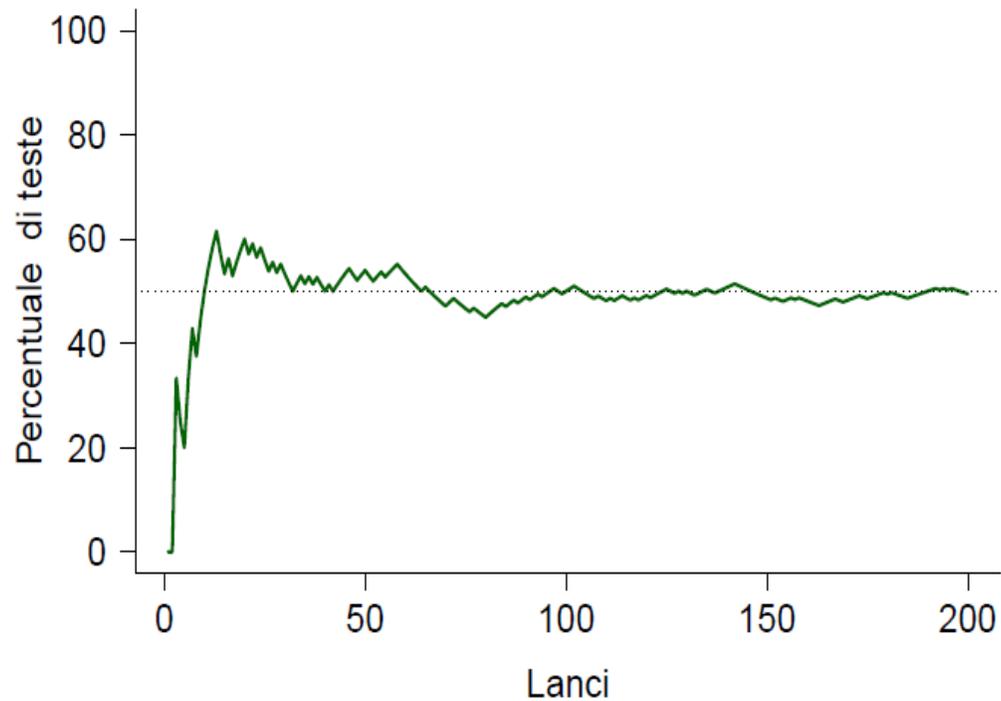
Per la moneta i casi possibili sono 2 (testa o croce) e quindi la probabilità che esca testa (o che esca croce) è pari ad $1/2$.



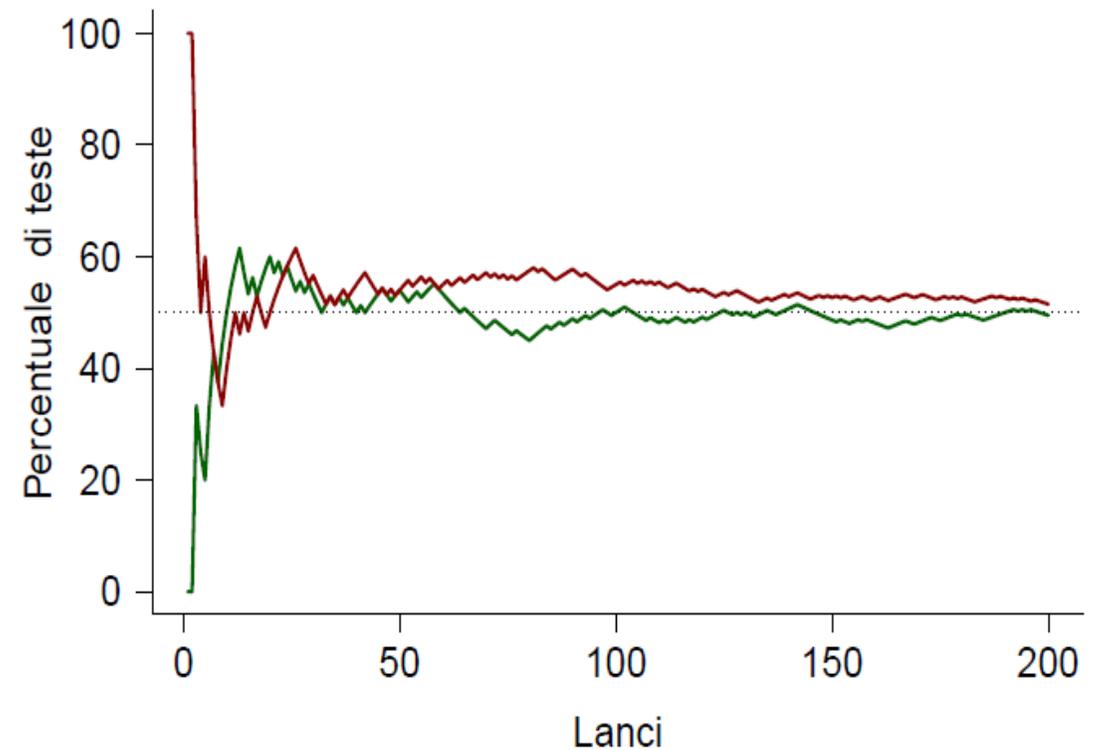
Limite “*concettuale*” di questa definizione:
per definire la probabilità occorre sapere preliminarmente che cosa significa che due casi sono ugualmente probabili, cioè sapere già che cosa è la probabilità!

All'aumentare dei lanci, la proporzione (*frequenza*) di un evento diventa sempre meno variabile e sempre più vicina ad un valore limite: tale proporzione 'di lungo periodo' viene definita *probabilità* dell'evento:

Questo è il risultato con 200 lanci



Facendo altri 200 lanci il risultato cambia



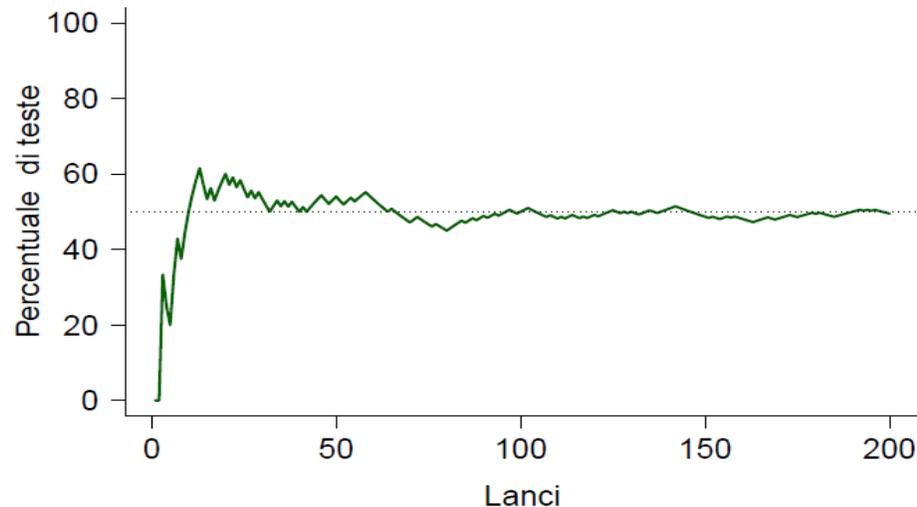
Man mano che si va avanti **il risultato si stabilizza intorno al 50%**, che è la probabilità che, intuitivamente, attribuiremmo all'evento 'esce testa'.

In una successione di prove fatte nelle stesse condizioni, la **frequenza** di un evento si avvicina alla probabilità dell'evento stesso, e l'approssimazione tende a migliorare con l'aumentare del numero delle prove*.

E quindi, secondo la definizione "*frequentista*" di probabilità:

La probabilità di un evento è il limite della frequenza (relativa) dei successi (cioè delle prove in cui l'evento si verifica) quando il numero delle prove tende all'infinito.

Questo è il risultato con 200 lanci

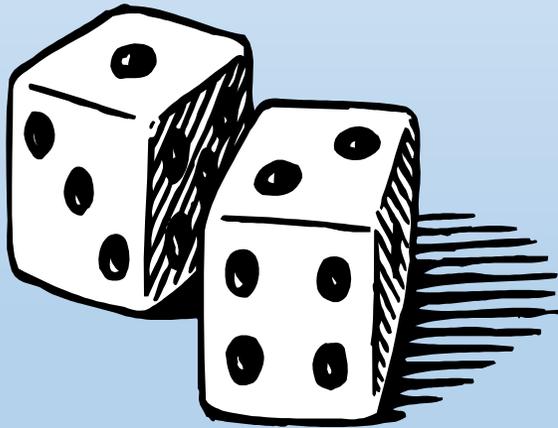


La probabilità dell'evento «esce Testa» corrisponde alla proporzione (frequenza relativa) della faccia testa: all'aumentare del numero dei lanci tale frequenza tende a $\frac{1}{2}$.

Anche questa definizione di probabilità è però alquanto criticabile: in generale, non potremo mai osservare una sequenza **infinita** di prove !!

..e se osserviamo solo una 'parte' di sequenza non possiamo affermare *con precisione* la probabilità di un certo evento, perchè la probabilità, come si è detto, è una proprietà **a lungo termine**...

... il concetto di sequenza '**infinita**' deve essere interpretato come una **cornice teorica ideale** (noi osserveremo sempre la frequenza degli eventi in **campioni di dimensione finita**).



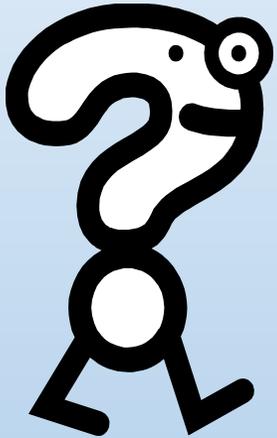
Se si lancia un dado la frequenza relativa di ogni singola faccia del dado è circa pari ad $1/6$ in una *lunga* *successione di lanci*;
in una coorte di neonati la frequenza del sesso maschile (o femminile) oscilla intorno a $1/2$...



La probabilità si misura con un numero tra 0 e 1:

- se un evento *non capita mai* - in nessuno dei *trials* della sequenza random- la sua probabilità è 0
- se un evento *capita sempre* - in tutti i *trials* della sequenza random- la sua probabilità è 1

Complichiamo un po' la faccenda...



Qual è la probabilità che il fumo contribuisca al cancro al polmone ?

Difficile rispondere in termini di una sequenza casuale di 'prove' in cui si osserva se in alcune il fumo è una causa di cancro al polmone ed in altre no...

Definizione “*soggettiva*”^{*} di probabilità:

La probabilità di un evento è il grado di fiducia - espresso tra 0 e 1 - che un individuo ha nel verificarsi dell'evento.

(Bruno De Finetti, 1906-1985)



In altri termini, e usando il linguaggio delle scommesse:

La probabilità di un evento è il prezzo che un individuo ritiene *equo* pagare per ricevere 1 se l'evento si verifica e 0 se l'evento non si verifica. Le probabilità degli eventi devono essere attribuite in modo tale che non sia possibile ottenere con un insieme di scommesse una vincita *certa* o una perdita *certa*.

Indipendentemente da quale sia la definizione di probabilità che usiamo, abbiamo bisogno di *strumenti matematici* per poterla utilizzare in pratica !!!



^{*} impostazione «**bayesiana**»

Cenni di calcolo delle probabilità

Le operazioni di base del calcolo delle probabilità sono l'addizione e la moltiplicazione.

Esempio: qual è la probabilità che esca 1 oppure 3 in un lancio di un dado?



Se il dado è bilanciato, la probabilità di 1 è $1/6$ e la probabilità di 3 è $1/6$.

In nessun lancio potranno uscire **contemporaneamente** 1 e 3.

L'evento "**composto**" definito come: «esce la faccia 1 oppure esce la faccia 3»:

$$P(\text{esce 1 oppure esce 3}) = 1/6 + 1/6 = 1/3.$$

La faccia 1 e la faccia 3 non possono uscire contemporaneamente: sono due eventi *mutualmente esclusivi (incompatibili)*.

Altrimenti, la **regola di addizione** non sarebbe stata valida.

Complichiamoci un po' la vita...

Supponiamo che il nome di un infermiere venga estratto **casualmente** da un registro, e qualcuno ci dica che la probabilità che l'infermiere sia maschio è di 0.4 (40%) mentre la probabilità si sia laureato a Trieste è di 0.8 (80%)...

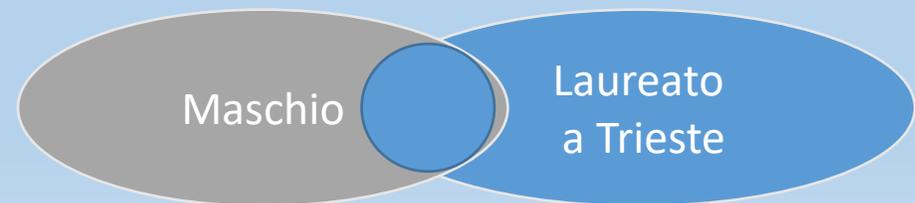
...qual è la probabilità che l'infermiere che abbiamo estratto sia maschio oppure si sia laureato a Trieste, oppure entrambe le cose?

Se le due probabilità venissero addizionate, il risultato sarebbe: $(0.4+0.8)=1.2$,

→ sbagliato poichè la probabilità di un evento, anche composto, non puo' mai superare 1 !!!

L'errore risiede nel fatto che la probabilità del doppio evento: «infermiere maschio e laureato a Trieste" è stata contata 2 volte, una volta come parte della probabilità di essere infermiere maschio e una volta come parte della probabilità di essere laureato a Trieste.

**i due eventi NON SONO mutualmente esclusivi
sono «compatibili»**



Indichiamo con:

P = probabilità dell'evento;

A = evento ['infermiere maschio'];

B=evento ['infermiere laureato a Trieste']

P(A e B) = probabilità dell'evento **congiunto** ['infermiere maschio e laureato a Trieste']

La forma piu' generale della *regola di addizione* è:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Supponiamo che la probabilità dell'evento **congiunto** P(A e B) è pari a 0.3; allora:

$$P(A \cup B) = (0.4 + 0.8) - 0.3 = 0.9.$$

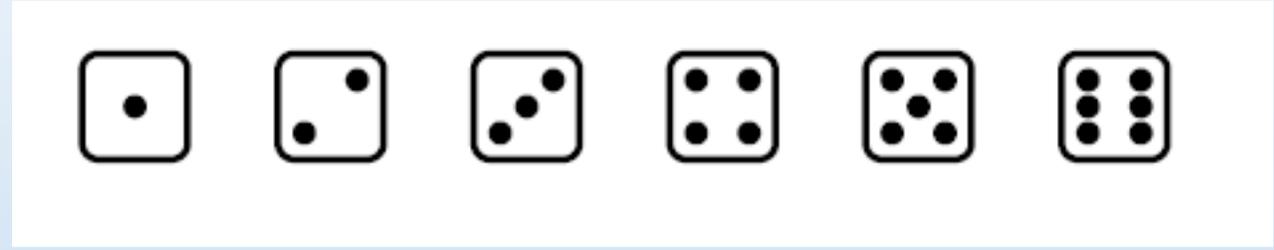
Se i due eventi A e B sono **mutualmente esclusivi**, allora:

P(A e B)=0, e così si ottiene la *forma semplice* della regola di addizione:

$$P(A \cup B) = P(A) + P(B)$$



Supponiamo adesso che ad ogni prova vengano lanciati **contemporaneamente** una moneta ed un dado: qual è la *probabilità congiunta* di testa (T) sulla moneta e 5 come faccia sul dado?



Regola di moltiplicazione:

$$P(T \text{ e } 5) = P(T) * P(5, \text{ dato testa}) = P(5) * P(T, \text{ dato } 5)$$

$$P(5, \text{ dato testa}) = \textit{‘probabilità condizionata’}$$

In questo particolare esempio, non c'è ragione di supporre che la probabilità di 5 sul dado sia in qualche modo 'condizionata' dall'evento Testa sulla moneta; in altre parole:

$$P(5, \text{ dato testa}) = P(5)$$

$$P(T, \text{ dato } 5) = P(T)$$

Notazione matematica: $P(A, \text{ dato } B) = P(A | B)$

Se la probabilità condizionata è uguale alla probabilità non condizionata, i due eventi sono definiti **indipendenti**, e si ottiene **la forma semplice** della regola di moltiplicazione:

$$P(T \text{ e } 5) = P(T) * P(5) = 1/2 * 1/6 = 1/12$$

Nell'esempio dell'infermiere maschio (A) e laureato a Trieste (B), se questi due eventi fossero indipendenti avremmo:

$$P(A \text{ e } B) = P(A) * P(B) = 0.4 * 0.8 = 0.3$$

Se questi due eventi invece non fossero indipendenti, perchè per esempio i corsi di laurea per infermieri a Trieste accettano come regola più donne che uomini, il valore corretto per $P(A \text{ e } B)$ andrebbe ottenuto accertando il valore della probabilità condizionata:

$$P(B=\text{laureato a Trieste} | A=\text{maschio})$$

Cenni di calcolo combinatorio: le permutazioni

Le permutazioni sono come gli anagrammi ed indicano **in quanti modi** (*ordinamenti*) **diversi** possono essere presi n oggetti.

Le permutazioni di n elementi sono date dallo *sviluppo fattoriale* di n (ossia n è moltiplicato per tutti i numeri interi inferiori ad n).

Ad esempio, date 4 lettere (ABCD), in quanti ordini si possono presentare?
(ABCD, BACD, DACB,)?

$$n! = n * (n - 1) * (n - 2) * (n - 3) * \dots$$

$$4! = 4 * 3 * 2 * 1 = 24$$

$$n! = \begin{cases} 1 & \text{se } n = 0 \\ n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1 & \text{se } n \in \mathbb{N}, n \neq 0 \end{cases}$$

Cenni di calcolo combinatorio: le disposizioni (I)

Disposizioni semplici:

Sono le possibili scelte di ***k* elementi ordinati** da un insieme composto da n oggetti (disposizioni di n elementi di classe k).

Ad esempio, quali sono i possibili podi (1,2,3) in una gara tra 8 atleti?

Al primo posto possono esserci 8 persone diverse, al secondo posto ce ne possono essere 7 e al terzo 6; le combinazioni possibili possono essere $8 \times 7 \times 6 = 336$.

In generale:

$$\text{Disposizioni} = \frac{n!}{(n - k)!}$$

$$\frac{8!}{(8-3)!} = \frac{8*7*6*5*4*3*2*1}{5*4*3*2*1} = 8*7*6 = 336$$

Le disposizioni coincidono con le permutazioni se $n = k$

Nota bene: **$0! = 1$**

Cenni di calcolo combinatorio: disposizioni (II)

Disposizioni con ripetizione:

Le disposizioni **con ripetizione** (disposizioni con ripetizione di n elementi di classe k) sono come le disposizioni, ma ogni oggetto, dopo essere stato scelto **viene rimesso** nell'insieme di partenza.

Date 10 lettere (da A a L), quante *combinazioni ordinate/disposizioni con ripetizione* da 4 lettere posso effettuare?

Si parla anche di *combinazioni ordinate* perchè l'ordine delle lettere conta, cioè ABCD, ad esempio, è diverso da BACD.

Per la prima lettera ho dieci possibilità, altrettante per la seconda e così via, cioè:

$$10 \times 10 \times 10 \times 10 = 10^4 = 10.000$$

In generale, le disposizioni con ripetizione di n elementi di classe k sono: n^k

Cenni di calcolo combinatorio: le combinazioni

Nelle combinazioni (di n elementi di classe k , con $k \leq n$) **non viene considerato l'ordine** con cui gli oggetti si presentano.

In quanti modi si possono selezionare k oggetti da un insieme di n quando non si tiene conto dell'ordine di estrazione?

$$\text{combinazioni} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Coefficiente binomiale

e per convenzione si ha che $\binom{n}{0} = 1$ e che vale la simmetria

$$\binom{n}{k} = \binom{n}{n-k}$$

Coefficiente binomiale

Il coefficiente binomiale rappresenta **il numero di sottoinsiemi** di k elementi che si possono estrarre da un insieme di n elementi.

Date le 10 (n) lettere da A ad L, le combinazioni a gruppi di 4 (k) che posso avere sono:

$$\frac{10!}{(10-4)!4!} = \frac{10!}{6!4!} = \frac{10*9*8*7*6*5*4*3*2*1}{6*5*4*3*2*1*4*3*2*1} = 10*3*7 = 210$$

In un trial clinico di tipo caso-controllo, su 10 controlli *eleggibili* se ne devono scegliere 5. Si hanno:

$$\binom{10}{5} = \frac{10 \times 9 \times 8 \times 7 \times 6}{5!} = 252$$

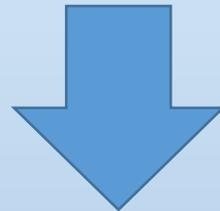
252 possibili campioni di 5 controlli.

Variabili Aleatorie e Distribuzioni di probabilità

Variabile aleatoria (VA): Prima di lanciare una moneta possiamo solo «predire» il risultato con una certa probabilità.

Il «lancio della moneta» è una variabile aleatoria.

il risultato di un esperimento che non può essere predetto con certezza (solo la probabilità di ottenere un certo risultato può essere stimata)



Le VA sono descritte da *distribuzioni di probabilità*

Es: Lancio di una moneta: testa o croce hanno probabilità $P=1/2$



Distribuzioni di Probabilità

Che cosa è una distribuzione di probabilità ?

E' una tabella (equazione) che ci dice quale è la probabilità di ciascun evento o risultato di un «esperimento»

Outcome X	Probability of outcome $P(X)$
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$

Questa tabella ci dice quale è la distribuzione di probabilità che otteniamo per i vari risultati dei lanci ripetuti di un dado.

L'equazione che genera questa distribuzione di probabilità è:

$$P(X)=1/6$$



Variabili casuali discrete

Le VA si dicono **DISCRETE** se il meccanismo che genera i dati è un conteggio oppure del tipo «presenza/assenza» (il *range di valori* è un insieme finito o numerabile).

Immaginiamo un mazzo di carte con i quattro assi, tre due, due tre e un quattro (10 carte) e definiamo la variabile casuale:

$$P(x = j) = \begin{cases} 0.4 & \text{se } x = 1 \\ 0.3 & \text{se } x = 2 \\ 0.2 & \text{se } x = 3 \\ 0.1 & \text{se } x = 4 \end{cases}$$

La funzione o *variabile casuale* $P(x)$ assegna una probabilità ad ogni carta; la somma delle probabilità di tutti gli eventi è sempre pari ad 1.

In questo caso i valori che la v.c. può assumere sono discreti e quindi si ottiene una distribuzione di probabilità discreta [una distribuzione di frequenza]

Si può definire anche la **funzione cumulata di frequenza**, detta anche **funzione di ripartizione**:

$$P(x \leq j) = \begin{cases} 0.4 & \text{se } x \leq 1 \\ 0.7 & \text{se } x \leq 2 \\ 0.9 & \text{se } x \leq 3 \\ 1.0 & \text{se } x \leq 4 \end{cases}$$

Possiamo definire poi la **media (valore atteso)** di una variabile casuale discreta come:

Expected Value: $\mu = E(X) = \sum x_i P(X = x_i)$

Possiamo infine definire anche la **varianza** di una variabile casuale discreta come:

$$\sigma^2 = Var(X) = E[X - E(X)]^2 = \sum (x_i - \mu)^2 P(X = x_i)$$

VA di Bernoulli

Jakob Bernoulli, 1654-1705



Una VA X di Bernoulli assume solo due valori: 0 o 1
fallimento/successo ; assenza/presenza.

La distribuzione di X è caratterizzata da una probabilità costante p di «successo» :

$$P(X=1)=p ; P(X=0)=1-p$$

Esempio: lancio di una moneta: $X=1$ se esce testa; $X=0$ esce croce $\rightarrow p(X=1)=p(X=0)=1/2$

$$X \approx \textit{Bernoulli}(p)$$

Una sequenza di lanci di una moneta è una sequenza di realizzazioni di una VA di Bernoulli

VA Binomiale

Una VA X Binomiale «conta» il numero k dei successi di n VA di Bernoulli.

$$Y = X_1 + X_2 + \dots + X_n \qquad Y \approx \text{Bin}(n, p)$$

Es: Quale è la probabilità che in n lanci di una moneta esca k volte testa?

Due parametri descrivono la distribuzione di X :

- il numero delle prove n
- la probabilità di successo p



Conta tutti i possibili modi in cui possiamo estrarre k elementi da n

→

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{1*2*3\dots*k} = \frac{n!}{(n-k)!k!}$$

Coefficiente binomiale:

$$P(Y = k) = \frac{n!}{(n-k)!k!} p^k q^{n-k}$$



$$E(X) = \sum_{i=1}^k x_i * P(x_i) \quad \text{Valore atteso (media) di una v.c. discreta}$$

Mean of binomial. The mean of the *binomial*(n, π) distribution is the sample size times the probability of success since

$$\begin{aligned} E[Y|\pi] &= \sum_{y=0}^n y \times f(y|\pi) \\ &= \sum_{y=0}^n y \times \binom{n}{y} \pi^y (1 - \pi)^{n-y} . \end{aligned}$$

We write this as a conditional mean because it is the mean of Y given the value of the parameter π . The first term in the sum is 0, so we can start the sum at $y = 1$. We cancel y in the remaining terms, and factor out $n\pi$. This gives

$$E[Y|\pi] = \sum_{y=1}^n n\pi \binom{n-1}{y-1} \pi^{y-1} (1 - \pi)^{n-y} .$$

Factoring $n\pi$ out of the sum and substituting $n' = n - 1$ and $y' = y - 1$, we get

$$E[Y|\pi] = n\pi \sum_{y'=0}^{n'} \binom{n'}{y'} \pi^{y'} (1 - \pi)^{n'-y'} .$$

We see the sum is a binomial probability function summed over all possible values. Hence it equals one, and the mean of the binomial is

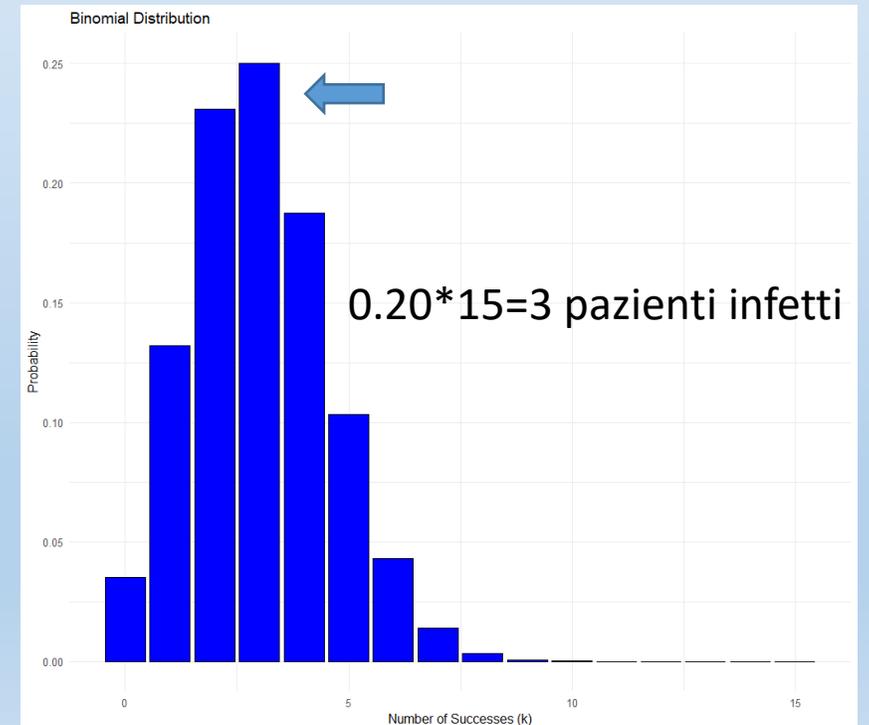
$$E[Y|\pi] = n\pi . \quad (5.7)$$

Il valore atteso della distribuzione di probabilità di una variabile casuale binomiale è il prodotto tra il numero di «prove» per la probabilità di successo in ogni prova.

La probabilità di contrarre una infezione in ospedale è una v.c. binomiale con $p=0.20$.

Abbiamo 15 pazienti ricoverati nel nostro reparto.

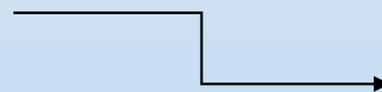
Quanti pazienti infetti «ci aspettiamo» di osservare (valore atteso?)



$$E(X) = \sum_{i=1}^k P(X = x_i)$$

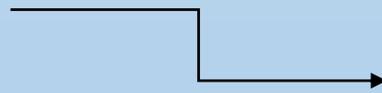
← Valore atteso (media) di una v.c. discreta

Valore atteso (media) della variabile binomiale:



$$E(X) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = np$$

Varianza della variabile binomiale:



$$\text{var}(X) = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq$$

Applicazioni della VA Binomiale

La distribuzione binomiale descrive il **numero** di volte in cui un particolare evento si verifica in una sequenza (casuale) di osservazioni (=campione).

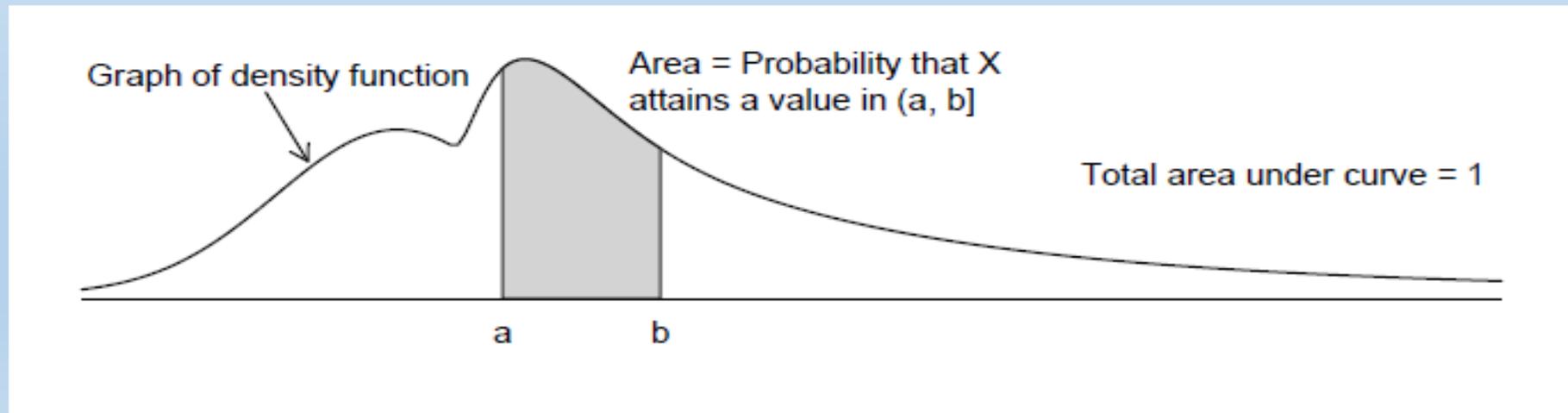
Questa distribuzione può essere utilizzata quando l'**outcome** di interesse è l'occorrenza di un evento, non la sua *"intensità"*.

- In un trial clinico, la condizione di un paziente può migliorare oppure no. Si intende valutare la probabilità di miglioramento, non *"di quanto"* i pazienti siano migliorati.
- Quante persone *"ansiose"* ci sono in questa aula ? La distribuzione binomiale potrebbe valutare il numero di persone ansiose, non il loro *"grado di ansia"*.
- Nel controllo di qualità, si vuole identificare il numero di lotti difettati in una filiera produttiva, *indipendentemente* dal tipo di difetto.

VA continua

- Le VA sono CONTINUE se il meccanismo generatore dei dati è un processo di misurazione su una scala numerica: peso corporeo, glicemia...
- Essendo la scala di misura continua ha senso chiedersi con quale probabilità i valori cadano in un certo **intervallo** più che identificare dei valori «puntuali»*...

Distribuzione di probabilità di una VA X continua: viene descritta tramite una “*funzione di densità*” la cui area sotto la curva rappresenta la probabilità:



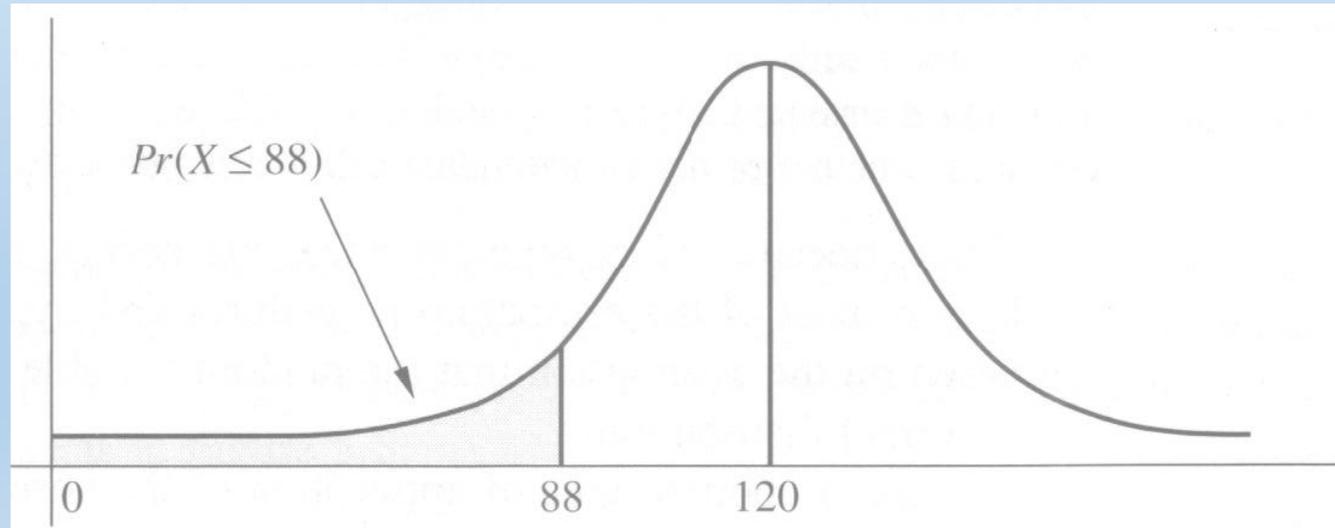
*essendoci potenzialmente infiniti valori...

La distribuzione di probabilità cumulativa di una variabile casuale continua X valutata in un punto a è pari alla probabilità che X abbia valori $\leq a$.

Rappresenta l'area sotto la curva 'alla sinistra' del valore a

Es: la distribuzione cumulativa del peso alla nascita valutata in 88 once (~2,5 kg) è pari a $P(X \leq 88)$ ed corrisponde all'area sotto la curva a sinistra di 88 once.

Questo valore è usato come limite per possibili outcomes sfavorevoli nel primo anno di vita:



Per una variabile casuale *continua* (dove ciò che si ottiene è detto *funzione di densità* o *densità di frequenza*) la funzione di ripartizione (=probabilità cumulata), la media e la devianza sono definite ricorrendo al concetto di *integrale*:

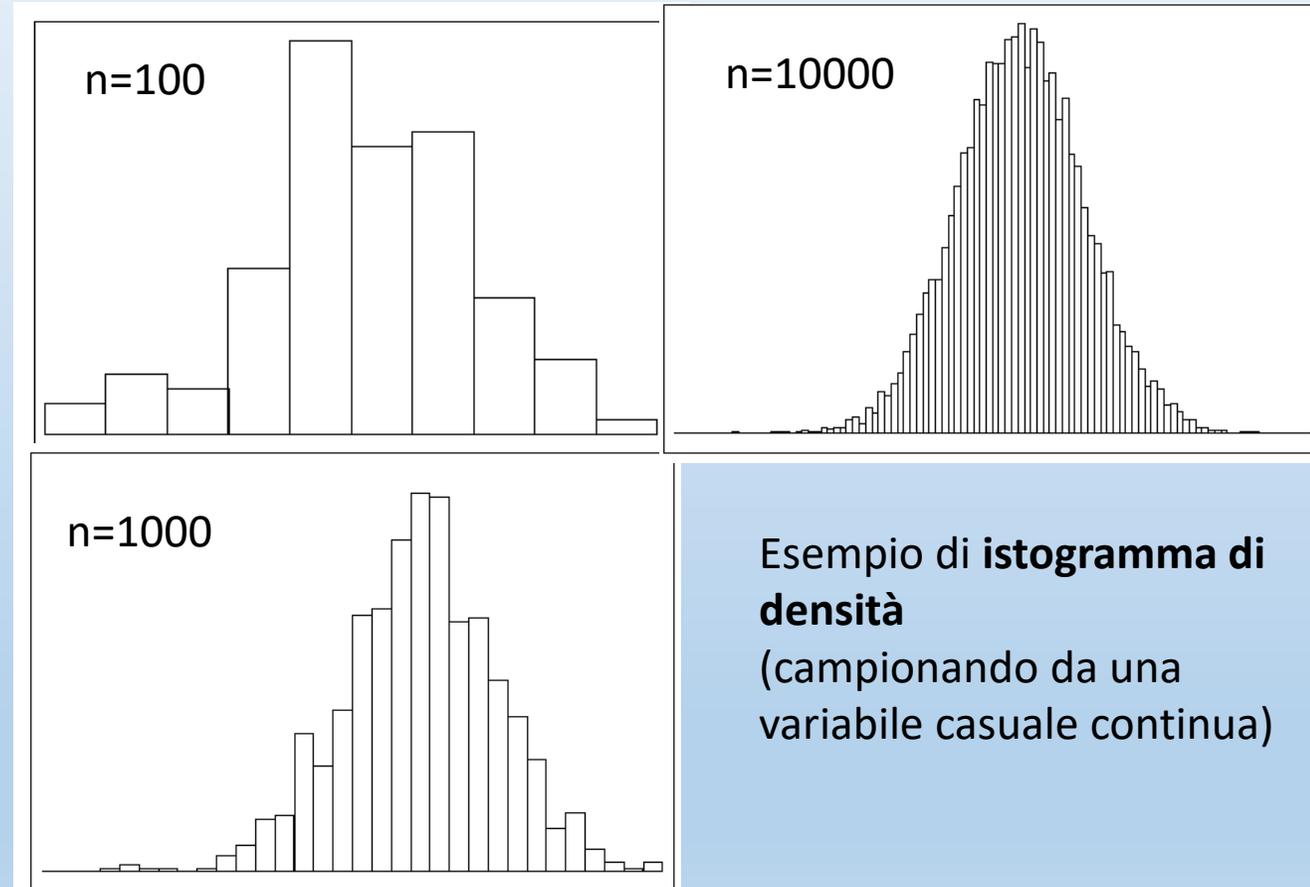
Funzione di densità $P(X) = f(x) \quad \int_{-\infty}^{\infty} f(x)dx = 1$

Funzione di ripartizione $P(X < x) = \int_{-\infty}^x f(x)dx$

Media o valore atteso $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$

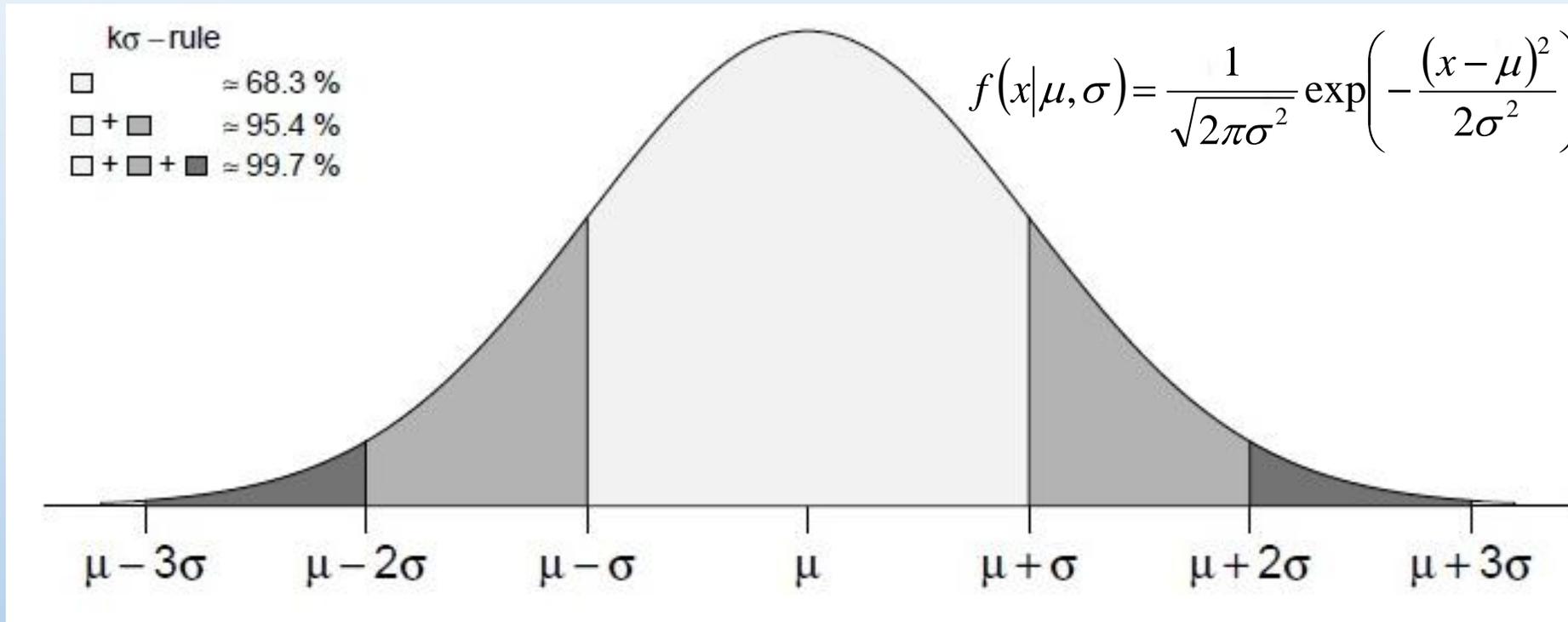
Varianza $\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

$$P(a < X < b) = \int_a^b f(x)dx$$



VA GAUSSIANA (o Normale)

Alcuni fenomeni (continui) possono essere descritti da una distribuzione di probabilità a forma di «campana»:



C.F. Gauss (1777-1855)

$N(\mu, \sigma^2)$ è definita dai parametri μ (=media di popolazione) e σ (=dev. standard di popolazione)

$N(0,1)$ distribuzione normale «standardizzata» ($\mu=0$ e $\sigma=1$) $\longrightarrow \frac{X - \mu}{\sigma} = Z \approx N(0,1)$

[OPZIONALE]

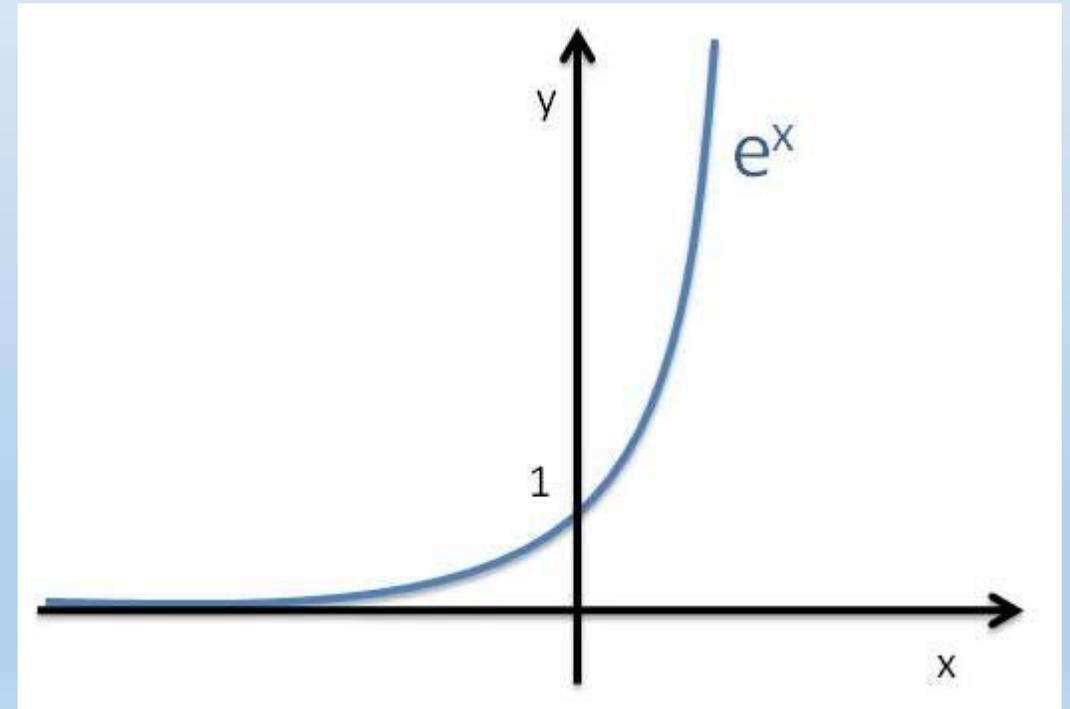
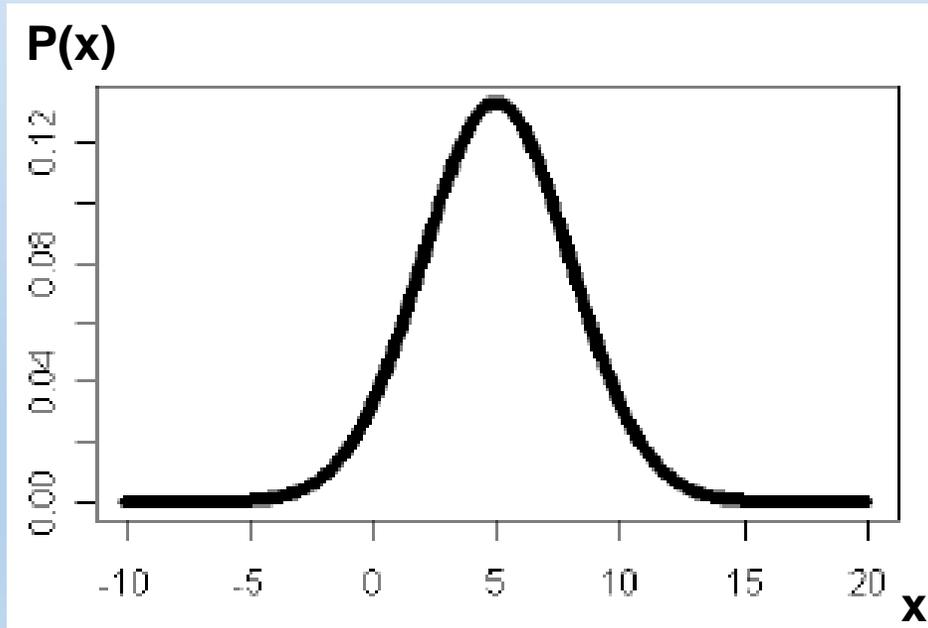
Densità della V.C. Gaussiana: **exp=e**

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

$\pi=3.14$

In matematica, la funzione esponenziale è la funzione che associa a un valore x l'elevamento a potenza con base il numero di Eulero e (una costante matematica il cui valore approssimato è 2.718) e ad esponente x . Viene solitamente rappresentata come e^x , oppure $\exp(x)$.

$N(\mu;\sigma)=N(5;3)$



Funzione di densità della V.C. Gaussiana standard: $\mathbf{N}(\mu;\sigma)=\mathbf{N}(0;1)$

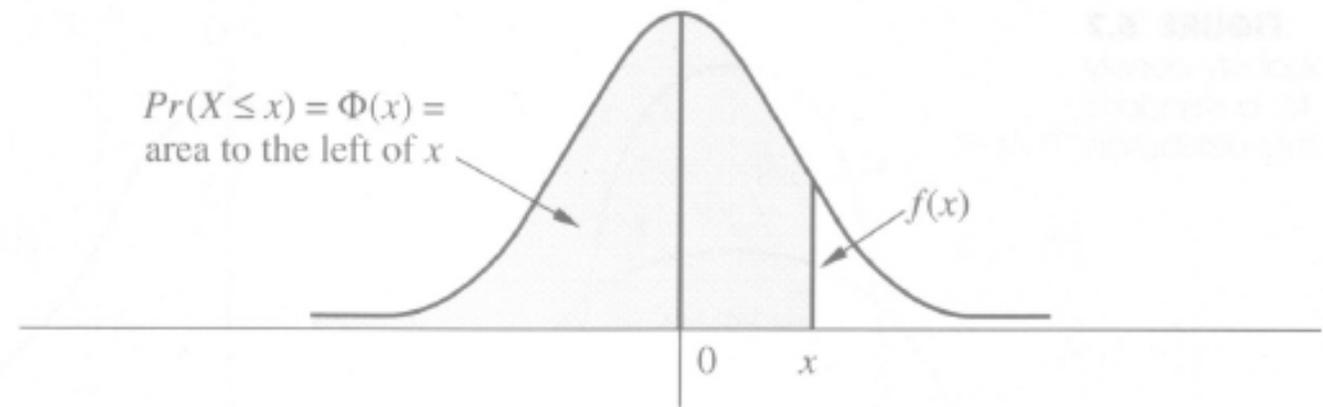
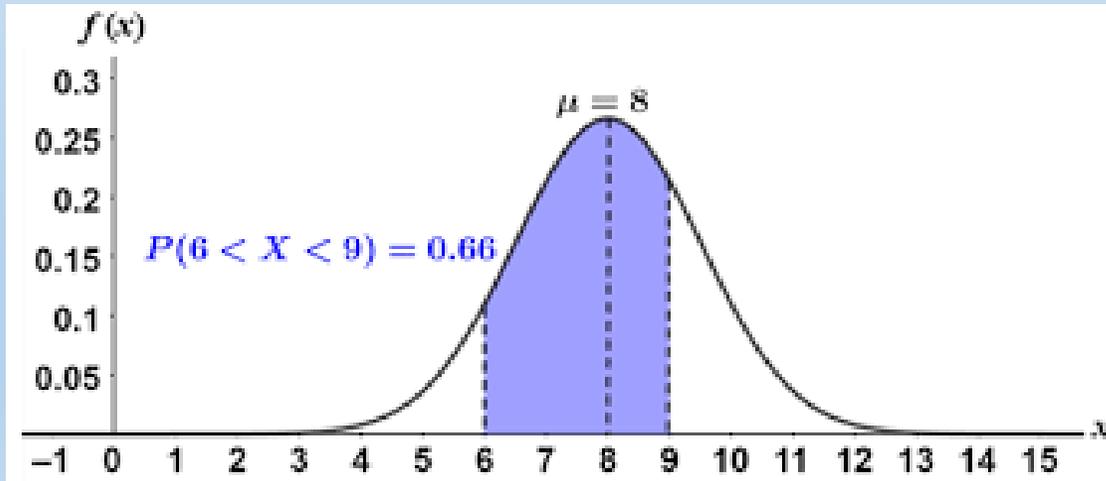
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < x < \infty$$

Possiamo definire **la probabilità di qualsiasi intervallo** utilizzando il calcolo integrale basato sulla distribuzione standardizzata.

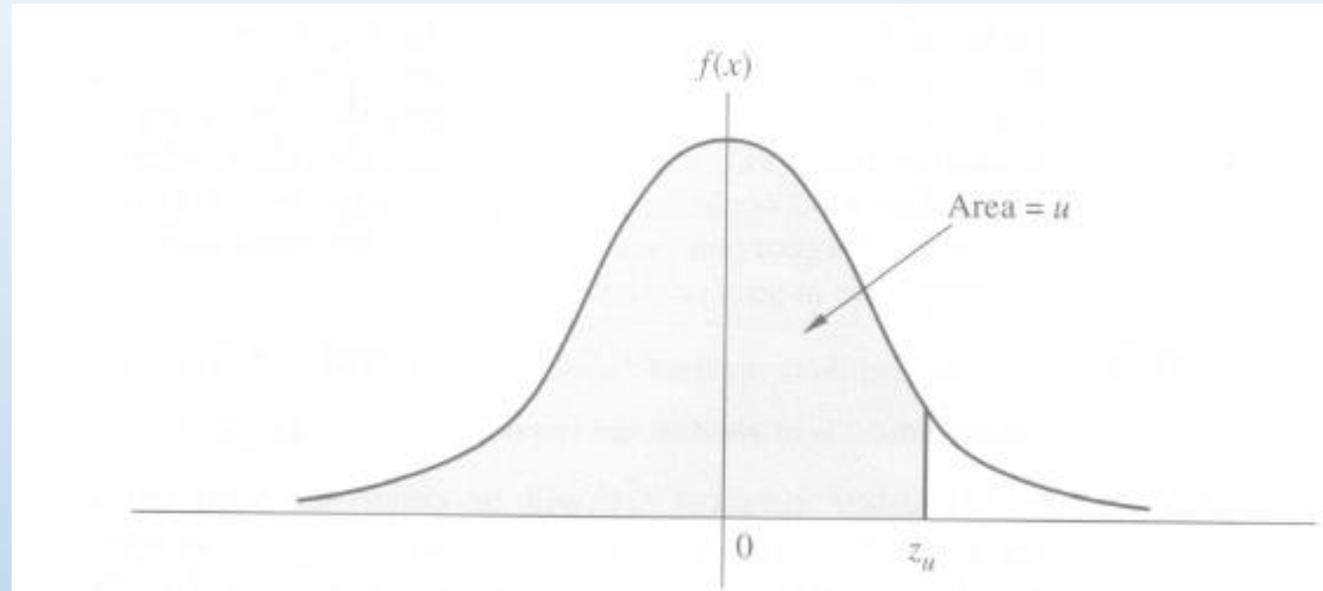
Funzione di ripartizione



$$\Phi(x) = P(X \leq x)$$



Calcolo dei *percentili* di una Normale standard



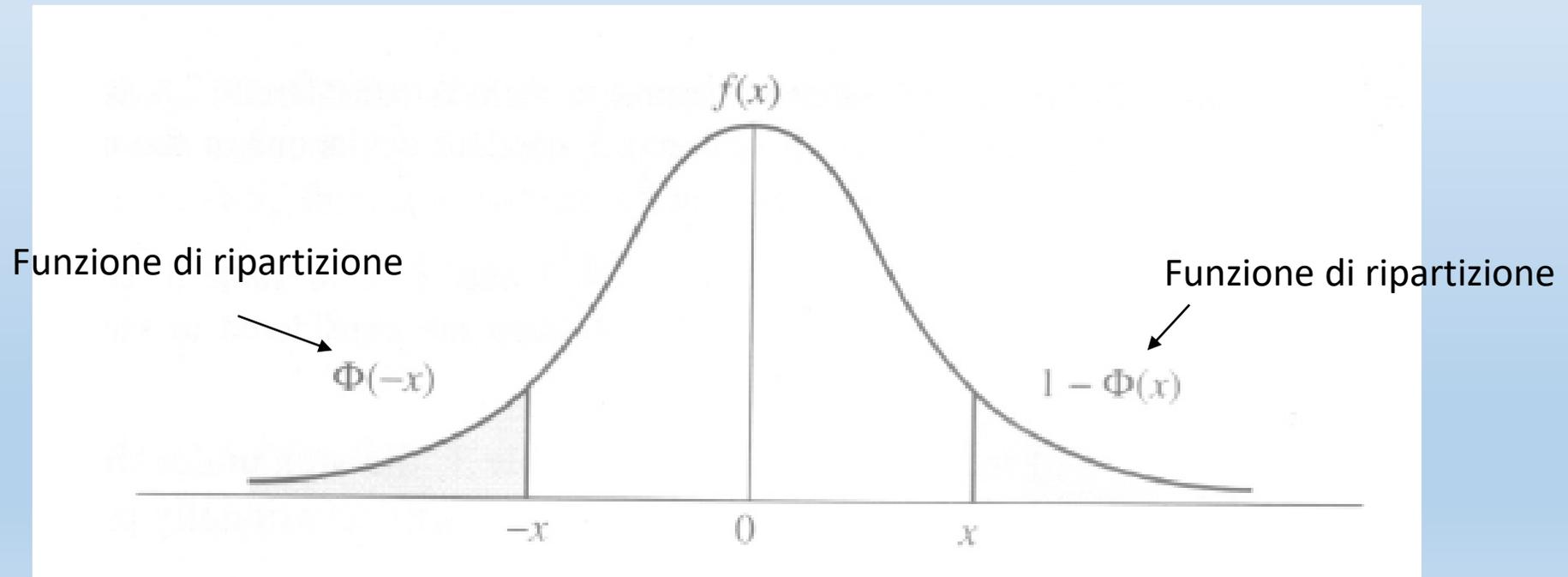
Il $(100-u)$ percentile di una normale standard è indicato con z_u ed è definito dalla relazione:

$$P(X < z_u) = u$$

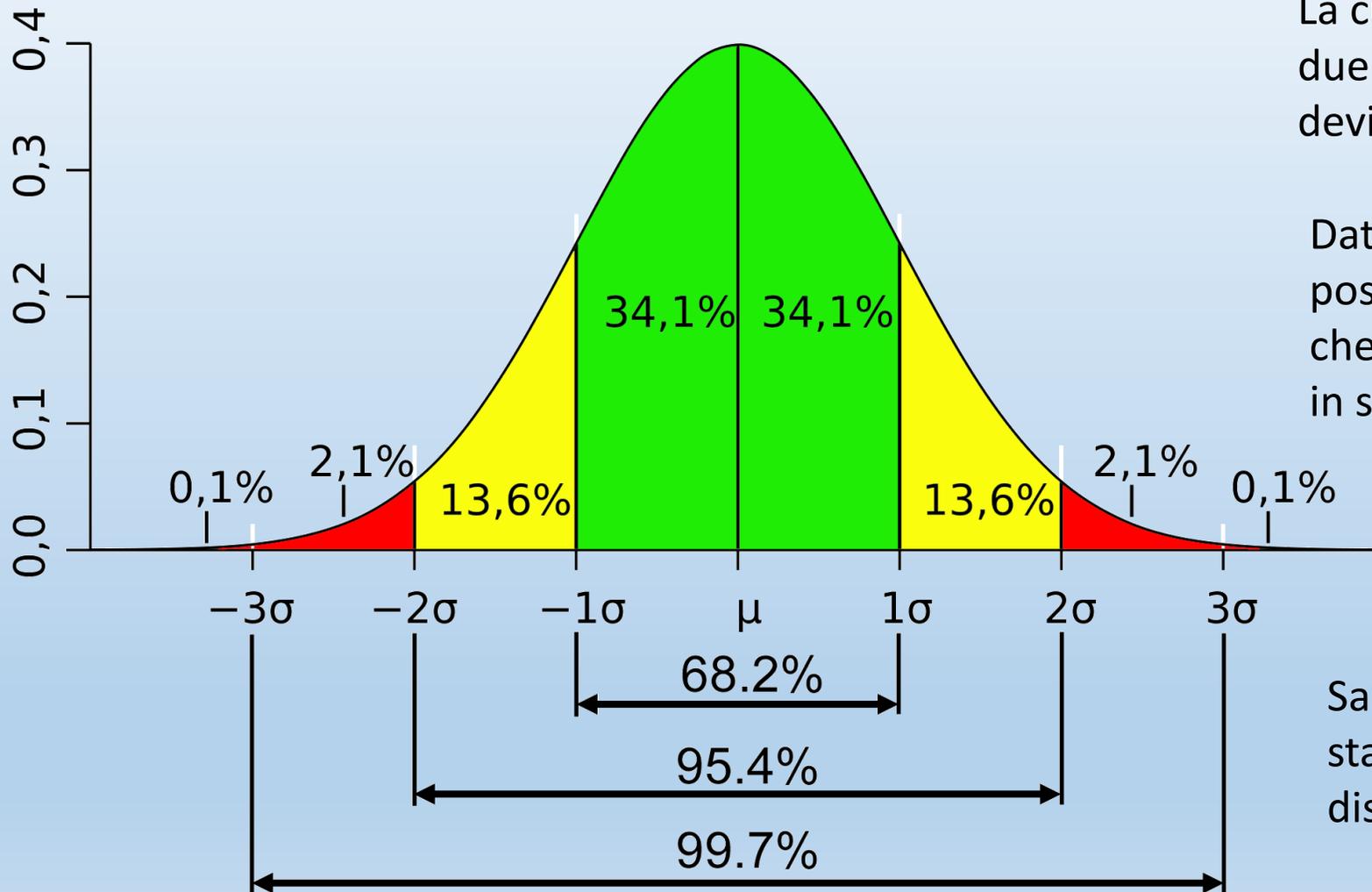
Per trovare z_u si deve trovare l'area u e quindi trovare il valore che corrisponde a tale area.

Si può usare la **simmetria** della distribuzione: $z_u = -z_{1-u}$

$$z_{.975} = 1.96 \quad z_{.95} = 1.645 \quad z_{.5} = 0 \quad z_{.025} = -1.96$$



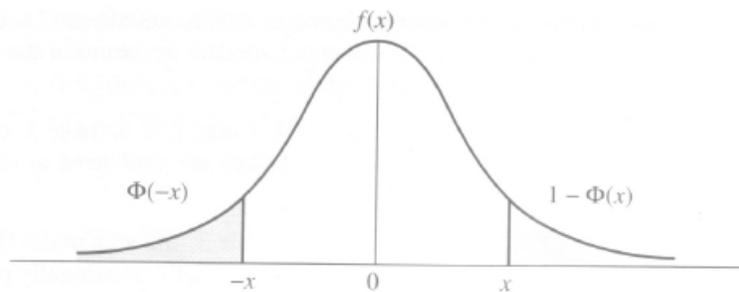
La «forma» della distribuzione normale (curva gaussiana)



La curva normale è caratterizzata da due parametri: la media μ e la deviazione standard σ .

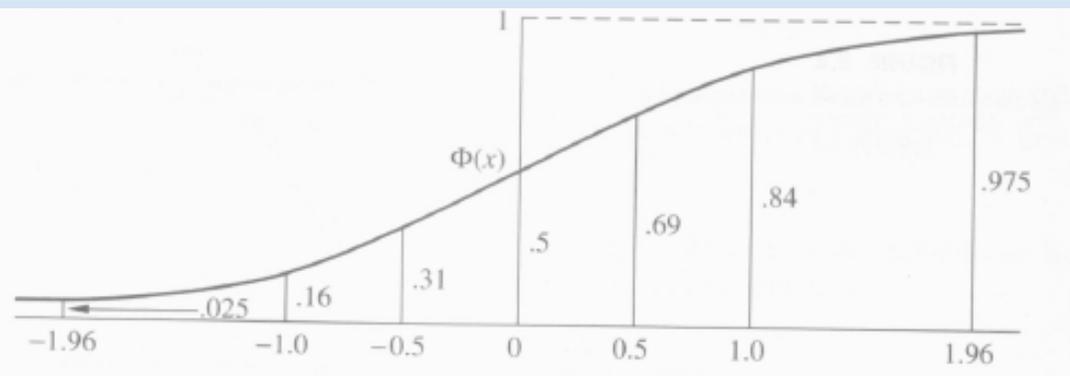
Dato il valore di questi due parametri, possiamo calcolare la % di unità statistiche che cadono in certe regioni della variabile in studio.

Sappiamo quindi anche la % di unità statistiche che cadono nelle «code» della distribuzione.



$$\Phi(-x) = P(X \leq -x) = P(X \geq x) = 1 - P(X \leq x) = 1 - \Phi(x)$$

Rappresentazione Cumulativa:



Si calcoli $P(X \leq 1.96)$ e $P(X \leq 1)$. Dalle tavole si ha che

$$\Phi(1.96) = 0.975 \quad \Phi(1) = 0.8413$$

Si calcoli $P(X \leq -1.96)$. Dalle tavole si ha che

$$\Phi(-1.96) = P(X \geq 1.96) = 0.0250$$

Ricapitolando:

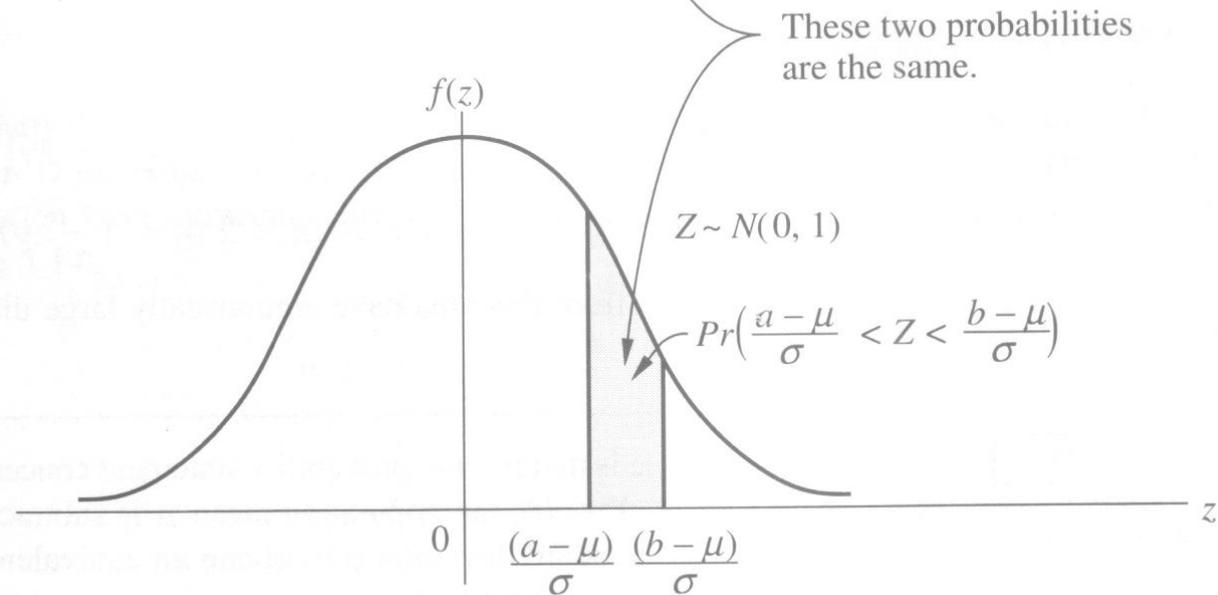
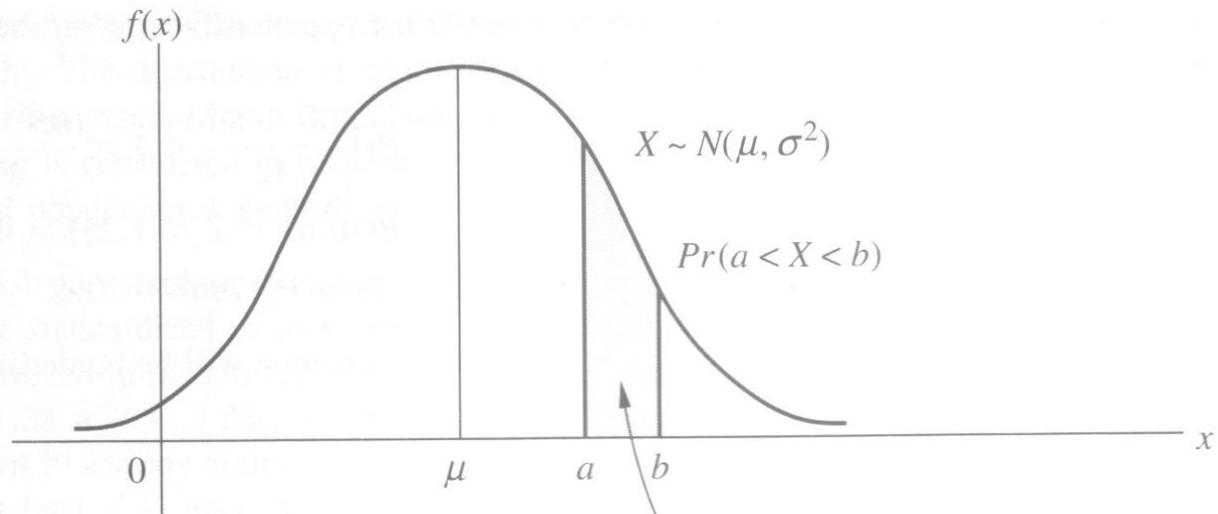
Dalla Normale alla Normale Standard:

Se: $X \sim N(\mu, \sigma^2)$ e $Z = \frac{X - \mu}{\sigma}$

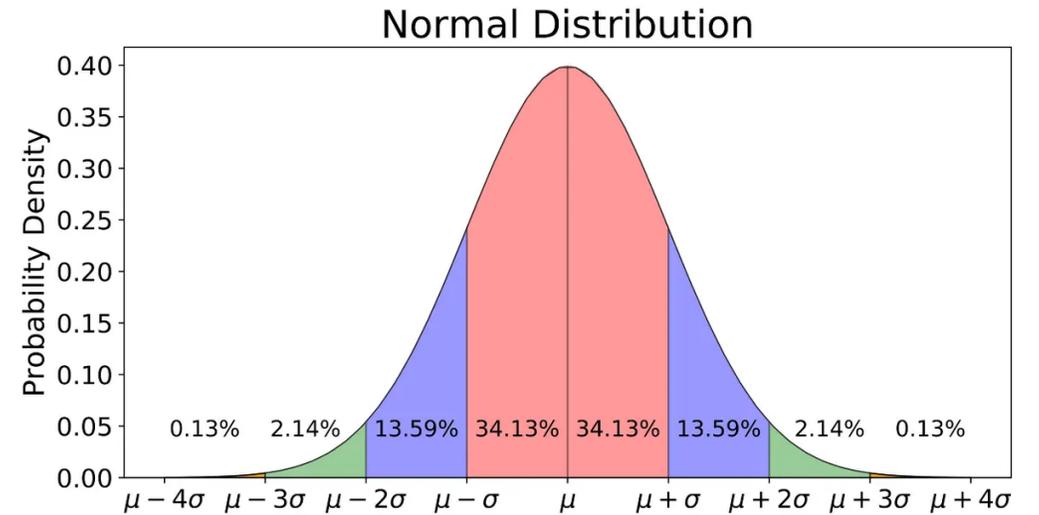
Allora: $Z \sim N(0, 1)$

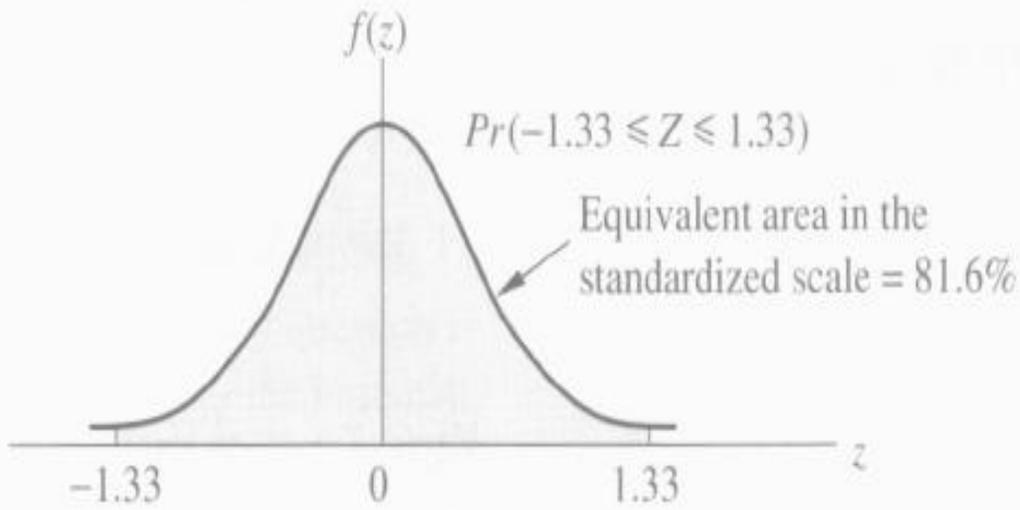
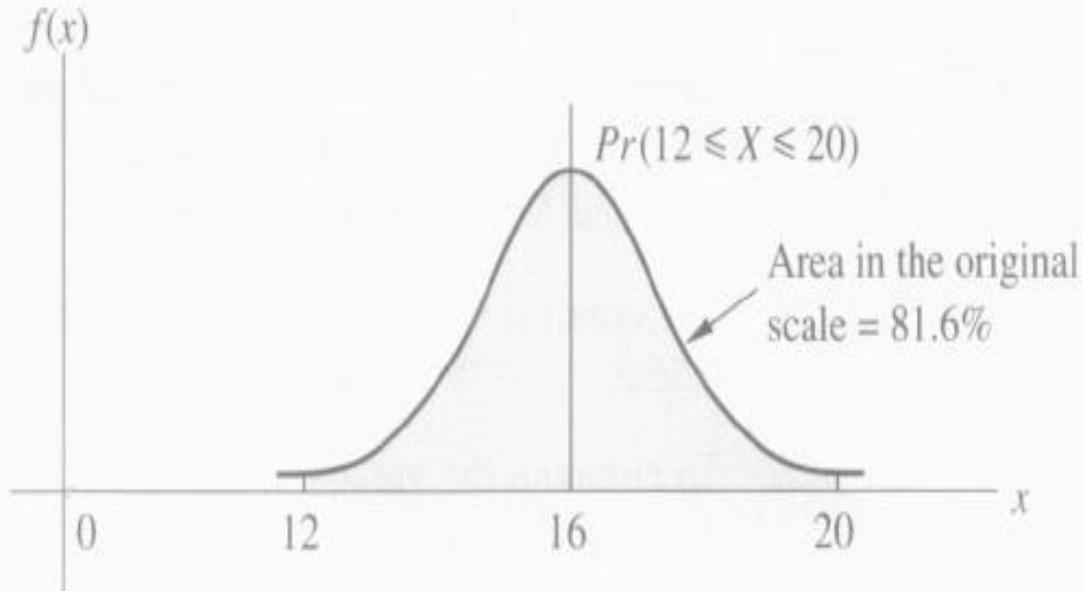
in tali condizioni si ha che:

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi[(b - \mu)/\sigma] - \Phi[(a - \mu)/\sigma]$$



Il motivo per cui Gauss ha deciso di utilizzare la v.c. Normale Standardizzata è legato al fatto che così gli integrali da risolvere per determinare le aree di probabilità in certi intervalli sono generali e non dipendono da μ da σ , valori che possono invece variare da applicazione ad applicazione.





Le distribuzioni normali sono **infinite** (perchè infiniti sono i valori possibili per μ e σ), ma possono tutte essere ricondotte **ad una sola distribuzione** di riferimento con $\mu=0$ e $\sigma=1$: la distribuzione normale standardizzata.

-> permette di risolvere il problema del calcolo di frequenza o di probabilità semplicemente ricorrendo alle tavole degli integrali della distribuzione normale standardizzata.

Una persona **borderline** per ipertensione è definita come una DBP compresa tra 90 e 95 mm Hg.
Nei maschi tra i 35 ed i 44 anni la DBP è **distribuita normalmente** con media 80 e varianza 144 (dev std=12).
Quale è la probabilità di essere borderline per un maschio in quella classe di età ??

$$\begin{aligned} P(90 < X < 95) &= P\left(\frac{90 - 80}{12} \leq Z \leq \frac{95 - 80}{12}\right) \\ &= \Phi[1.25] - \Phi[0.83] = 0.8944 - 0.7967 = 0.098 \end{aligned}$$

La probabilità di essere borderline per un maschio tra 35 e 44 anni è pari a 0.098 (circa 10%)

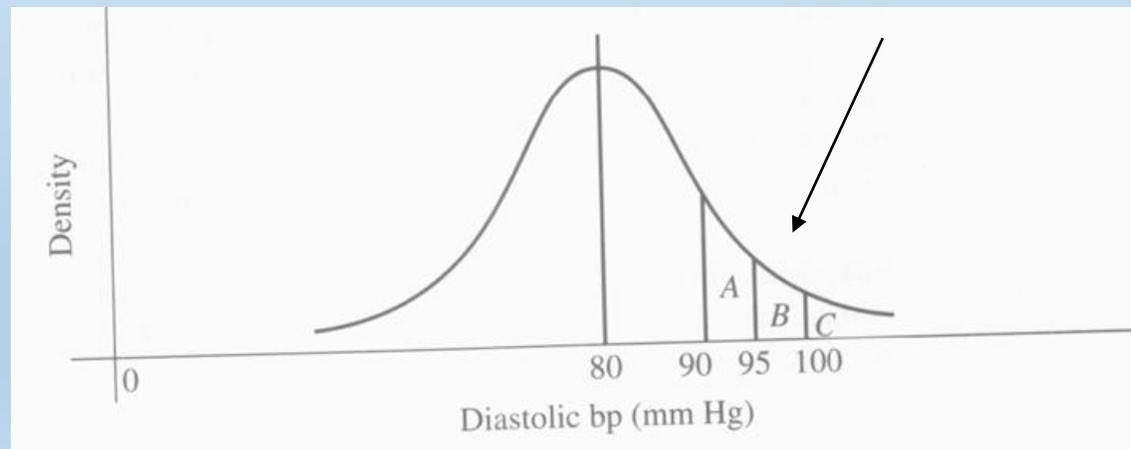
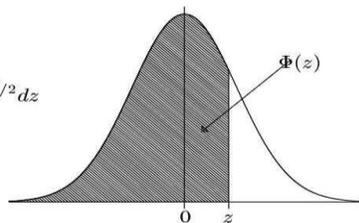


Tavola 1: Funzione di ripartizione della Variabile Casuale Normale Standardizzata

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



Oggi non usiamo più le tavole di Gauss!

Ci viene in aiuto il computer, e i software di statistica !!

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

1- 0.1056=0.8944

1- 0.2033=0.7967

$\Phi(1.25) - \Phi(0.83) = 0.098$