

LEZIONI DI STATISTICA MEDICA

Prof. SIMONE ACCORDINI

Lezione n.9

- Teoria della probabilità



Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

ELEMENTI DI TEORIA DELLA PROBABILITA'

La TEORIA DELLA PROBABILITA' ci permette
di studiare e descrivere gli **eventi aleatori**.

DEFINIZIONE: un evento è **aleatorio** quando di esso
non si può predire con certezza il risultato.

Esempi:

- numero estratto al lotto
- faccia di una moneta
- presenza di un'infezione virale



SPAZI CAMPIONARI ED EVENTI

Un **esperimento** è un qualsiasi processo di osservazione o misurazione.

Esempi:

- estrazione di un numero al lotto
- lancio di una moneta
- valutazione della presenza di un'infezione virale



Ad ogni esperimento è associato uno **spazio campionario S** costituito dall'insieme dei possibili risultati.

I singoli risultati dell'esperimento sono detti **elementi di S** o **eventi elementari**.



E' possibile definire un qualsiasi evento A come combinazione di più eventi elementari → **EVENTO COMPOSTO**

**SPAZIO
CAMPIONARIO (S)**

EVENTO (A)

{TT, TC, CT, CC}

'almeno una testa nel
lancio di due monete'

= {TT, TC, CT}

{1, 2, 3, 4, 5, 6}

'numero dispari nel
lancio di un dado'

= {1, 3, 5}



Un evento A é un **SOTTOINSIEME** dello spazio campionario S:

$$A \subset S$$

Un evento è **CERTO**
se comprende tutti gli elementi di S



$$A = S$$

Un evento è **IMPOSSIBILE**
se non comprende alcun elemento di S



$$A = \emptyset$$



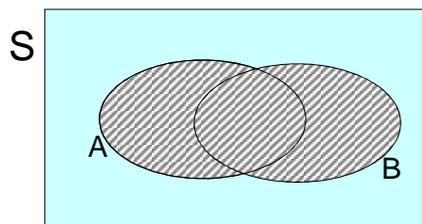
OPERAZIONI LOGICHE SUGLI EVENTI

Dato uno spazio campionario S e degli eventi $\{A, B, C, \dots\}$
in esso inclusi, è possibile definire nuovi eventi mediante
OPERAZIONI LOGICHE.



UNIONE DI EVENTI (somma logica): $A \cup B$

Siano A e B due eventi associati ad un esperimento:
l'evento C è definito **unione di A e B** se comprende
tutti gli elementi che appartengono ad A oppure a B.

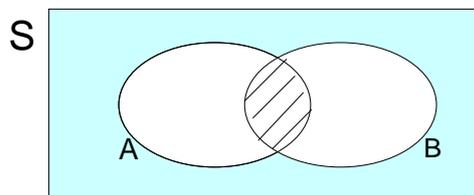


$$C = A \cup B$$



INTERSEZIONE DI EVENTI (prodotto logico): $A \cap B$

Siano A e B due eventi associati ad un esperimento:
l'evento C è definito **intersezione di A e B** se comprende
tutti gli elementi che appartengono ad A e a B.



$$C = A \cap B$$



Esempio: nel lancio di un dado sia

A = 'numero pari'

B = 'numero ≥ 4 '



Si determini l'insieme unione e l'insieme intersezione.

$$A = \{2, 4, 6\}$$

$$B = \{4, 5, 6\}$$

UNIONE



$$A \cup B = \{2, 4, 5, 6\}$$

INTERSEZIONE



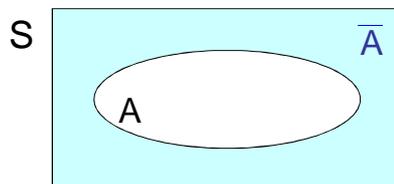
$$A \cap B = \{4, 6\}$$



NEGAZIONE DI UN EVENTO: \bar{A}

Dato un evento A, la sua negazione identifica un nuovo evento \bar{A} costituito da tutti gli elementi di S non appartenenti ad A.

\bar{A} è detto **complemento di A in S**.



Segue che:

$$A \cup \bar{A} = S$$

Se due eventi A e B non hanno elementi in comune, essi sono detti **EVENTI DISGIUNTI** o **MUTUAMENTE ESCLUSIVI** perché l'occorrenza dell'uno esclude l'altro.

Se A e B sono mutuamente esclusivi $\Rightarrow A \cap B = \emptyset$



PROBABILITÀ

Il concetto di probabilità ci permette di graduare l'ambito delle possibilità o di precisare il grado di fiducia che abbiamo nel verificarsi di un evento.



La **TEORIA DELLA PROBABILITA'** ci permette di **formulare delle valutazioni numeriche di probabilità** e di ricondurle alle regole del calcolo matematico.

L'**interpretazione** di tali valori numerici dipende dal significato che viene attribuito al concetto di probabilità.



CONCEZIONE CLASSICA DELLA PROBABILITÀ

La probabilità di un evento A è il rapporto tra il **numero di casi favorevoli** al verificarsi di A (n) e il **numero di casi possibili** (N)

$$P(A) = \frac{n}{N}$$

Esempi:

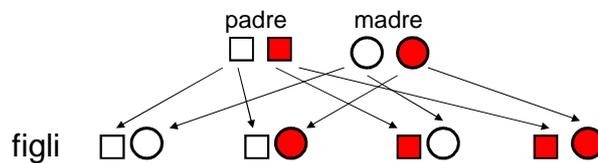
- probabilità di estrarre un asso da un mazzo di 52 carte = $4/52 = 0.08$
- probabilità di ottenere testa nel lancio di una moneta = $1/2 = 0.5$



- Tale definizione vale se i possibili risultati sono **equi-probabili** (gioco d'azzardo)
- Scarsamente applicabile in molte situazioni reali

Esempio di applicazione in medicina - Malattie genetiche

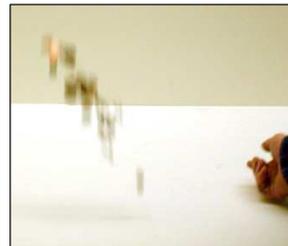
Se entrambi i genitori sono portatori sani del gene della talassemia o della fibrosi cistica, la probabilità di avere un figlio malato è 0.25.



CONCEZIONE FREQUENTISTA DELLA PROBABILITÀ

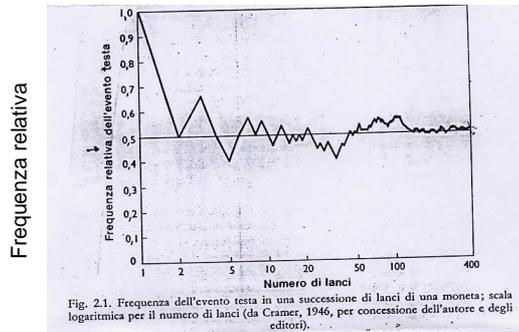
La probabilità di un evento A è la **frequenza relativa di successo** (occorrenza di A) in una serie di prove tendenti all'infinito, ripetute sotto identiche condizioni:

$$P(A) = \lim_{N \rightarrow \infty} \frac{n}{N}$$



Nel caso della concezione frequentista la probabilità viene assegnata sulla base dei risultati di un esperimento ripetuto molte volte (es. 1) nelle stesse condizioni o sulla base di situazioni che possono essere ricondotte a tale contesto concettuale (ad esempio: utilizzo di statistiche correnti, es. 2).

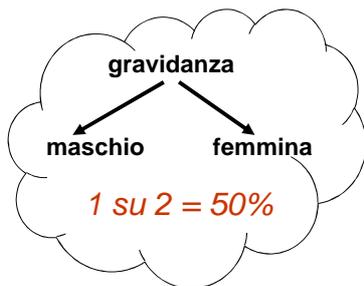
Esempio 1:
Frequenza dell'evento testa in una successione di lanci di una moneta.



Esempio 2:
Prob. che un bambino italiano nasca morto nel 2000 = $\frac{\text{numero nati morti nel 2000}}{\text{numero nati nel 2000}}$



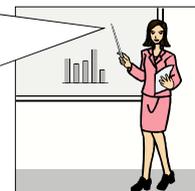
ESEMPIO: QUAL È LA PROBABILITÀ CHE UN NEONATO SIA FEMMINA?



(definizione **CLASSICA** di probabilità)

Però nel mondo, in assenza di interventi dell'uomo (aborti o infanticidi selettivi, omessa denuncia) nascono 1057 maschi ogni 1000 femmine.

$$1000 / (1000+1057) = 48,6\%$$



(definizione **FREQUENTISTA** di probabilità)



TEORIA ASSIOMATICA DELLA PROBABILITA'

Qualsiasi sia la definizione di probabilità, per probabilità (P) si intende una funzione a valori reali definita sullo spazio campionario S che soddisfa i seguenti assiomi:

1) per qualsiasi evento $A \subset S$, si ha che:

$$0 \leq P(A) \leq 1$$

in particolare: $P(A) = 1$ se A è l'evento **certo**

$P(A) = 0$ se A è l'evento **impossibile**

2) $P(S) = 1$



3) se $\{A_1, A_2, \dots, A_i, \dots\}$ sono una sequenza finita o infinita di eventi mutuamente esclusivi (o disgiunti) di S, allora

$$P(A_1 \cup A_2 \cup \dots \cup A_i \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_i) + \dots$$

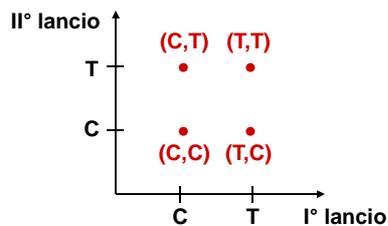
Il modo più elementare per assegnare una funzione di probabilità allo spazio campionario S è quello di assegnare una probabilità ad ogni elemento di S

\Rightarrow la probabilità corrispondente ad un qualsiasi **evento composto A** sarà definita come somma delle probabilità degli eventi elementari contenuti in A (*assioma 3*).

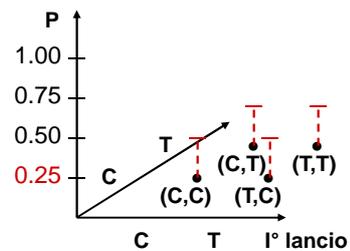


Esempio: Si consideri lo spazio campionario S associato al lancio di una moneta per due volte consecutive. In base alla definizione classica di probabilità, possiamo attribuire ad ogni punto dello spazio campionario S una probabilità $P = 1/4 = 0.25$.

SPAZIO CAMPIONARIO



FUNZIONE DI PROBABILITA'



Probabilità di ottenere almeno una testa

$$= P\{(T,C) \cup (C,T) \cup (T,T)\}$$

$$= P(T,C) + P(C,T) + P(T,T) = 0.25 + 0.25 + 0.25 = \mathbf{0.75}$$



MAZZO DI 52 CARTE

ESERCIZIO:

1. calcolare la probabilità di estrarre l'asso di cuori

$$\mathbf{1 / 52 = 0.02}$$

2. calcolare la probabilità di estrarre una carta rossa

$$\mathbf{26 / 52 = 0.5}$$

3. calcolare la probabilità di estrarre una figura

$$\mathbf{12 / 52 = 0.23}$$

4. calcolare la probabilità di estrarre un re

$$\mathbf{4 / 52 = 0.08}$$



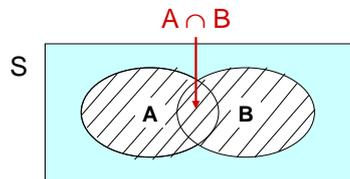
REGOLE DEL CALCOLO DELLA PROBABILITÀ

Il calcolo della probabilità è estremamente utile per stabilire sia la probabilità associata ad un evento, sia la probabilità associata ad un insieme di eventi.

REGOLA DELL'ADDIZIONE

Se A e B sono due eventi in S tali che $A \cap B \neq \emptyset$ (**eventi non disgiunti**):

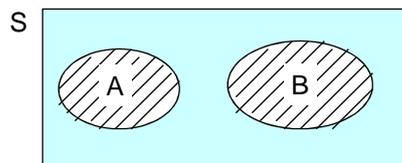
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



REGOLA DELL'ADDIZIONE

Se A e B sono due eventi in S tali che $A \cap B = \emptyset$ (**eventi disgiunti**):

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



Esercizio: calcolare la probabilità di estrarre una carta rossa o una figura da un mazzo di 52 carte (**eventi non disgiunti**)

$$P(\text{carta rossa}) = 26 / 52 = 0.5$$

$$P(\text{figura}) = 12 / 52 = 0.23$$

$$P(\text{carta rossa} \cap \text{figura}) = 6 / 52 = 0.11$$

$$\begin{aligned} P(\text{carta rossa} \cup \text{figura}) \\ = 0.5 + 0.23 - 0.11 = 0.62 \end{aligned}$$

Esercizio: calcolare la probabilità di estrarre una figura o una carta compresa tra 3 e 6 da un mazzo di 52 carte (**eventi disgiunti**)

$$P(\text{figura}) = 12 / 52 = 0.23$$

$$P(\text{carta } 3\div 6) = 16 / 52 = 0.31$$

$$P(\text{figura} \cap \text{carta } 3\div 6) = 0 !!!$$

$$\begin{aligned} P(\text{figura} \cup \text{carta } 3\div 6) \\ = 0.23 + 0.31 = 0.54 \end{aligned}$$



Se \bar{A} è il complemento di A in S:

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$$

$$P(\bar{A}) = 1 - P(A)$$

Esercizio: se la probabilità di morire nel 1° anno dalla diagnosi per un paziente affetto da tumore al polmone è pari a 0.30, qual è la probabilità di sopravvivere al 1° anno?

$$P(\text{sopravvivere}) = 1 - 0.30 = 0.70$$



PROBABILITÀ CONDIZIONALE E REGOLA DELLA MOLTIPLICAZIONE

Talvolta è molto utile conoscere la probabilità di un evento A in S quando si è verificato un altro evento B in S

→ **PROBABILITÀ CONDIZIONALE**

Esempio:

probabilità di uscita del 7 di quadri dato che è uscita una carta di quadri

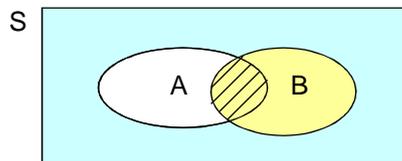
probabilità di avere un tumore al polmone dato che si fuma

probabilità di avere il colera data la presenza di una gastroenterite acuta



Se A e B sono due eventi dello spazio campionario S, si definisce **probabilità condizionale di A dato B**:

$$P(A | B) = P(A \cap B) / P(B)$$



N.B.: lo spazio campionario dell'evento B diviene il nuovo spazio campionario.



Dalla definizione di probabilità condizionale segue la **REGOLA DELLA MOLTIPLICAZIONE**:

$$P(A \cap B) = P(B) \cdot P(A | B)$$

$$P(A \cap B) = P(A) \cdot P(B | A)$$

Se il verificarsi di B non condiziona la probabilità del verificarsi di A, segue che:

$$P(A | B) = P(A)$$

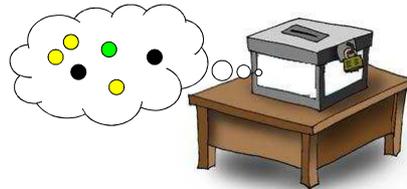
e i due **eventi** sono detti **indipendenti**, ovvero:

$$P(A \cap B) = P(A) \cdot P(B)$$



Esercizio:

Qual è la probabilità di estrarre senza reimbussolamento due palline gialle da un'urna che contiene tre palline gialle, due nere e una verde?



$$A = \text{estraz. I}^{\text{a}} \text{ pallina gialla} \quad \rightarrow \quad P(A) = 3/6$$

$$B = \text{estraz. II}^{\text{a}} \text{ pallina gialla} \quad \rightarrow \quad P(B | A) = 2/5$$

$$P(A \cap B) = P(A) \cdot P(B | A) = (3/6)(2/5) = \mathbf{1/5}$$

Se l'estrazione fosse con reimbussolamento:

$$P(B | A) = P(B) = P(A) = 3/6$$

$$P(A \cap B) = (3/6)(3/6) = \mathbf{1/4}$$



STIME DI PROBABILITA'

EVENTI DI MAGGIORE INTERESSE IN AMBITO MEDICO:

- malattia / morte (presente, M+; assente, M-)
- esposizione pregressa (presente, E+; assente, E-)

PROBABILITA' DI MAGGIORE INTERESSE IN AMBITO MEDICO (PROBABILITA' CONDIZIONALI):

- $P(M+ | E+)$
- $P(M+ | E-)$



esempio: la probabilità di morte per carcinoma polmonare tra gli individui di sesso maschile di età compresa tra i 55 e i 75 anni è:

$$P(M+) = 0.02$$

informazione quantitativa di
natura descrittiva

esempio: la probabilità di morte per carcinoma polmonare tra gli individui di sesso maschile di età compresa tra i 55 e i 75 anni **fumatori / non fumatori** è:

$$P(M+|E+) = 0.08$$

$$P(M+|E-) = 0.01$$

informazione quantitativa
sull'associazione tra
esposizione e malattia



Le probabilità condizionali e il rischio relativo richiedono la stima delle probabilità associate agli elementi dello SPAZIO CAMPIONARIO:

$$S = \{M+E+, M+E-, M-E+, M-E-\}$$

		MALATTIA	
		+	-
ESPOSIZIONE	+	$M+ \cap E+$	$M- \cap E+$
	-	$M+ \cap E-$	$M- \cap E-$



		MALATTIA	
		+	-
ESPOSIZIONE	+	$M^+ \cap E^+$	$M^- \cap E^+$
	-	$M^+ \cap E^-$	$M^- \cap E^-$

In assenza di informazioni a priori, le probabilità associate agli elementi dello spazio campionario possono essere stimate tramite le **frequenze con cui gli eventi elementari si sono verificati nel campione:**

$$P(.) = \frac{n(.)}{n}$$



Esempio: valutiamo la relazione tra allattamento al seno (E) e insorgenza di infezioni del primo tratto respiratorio nei primi 4 mesi dalla nascita (M)



Indagine condotta sui nati tra il 1982 e il 1983 in una clinica ostetrica dell'Arizona

		Infezione respiratoria		
		+	-	
Allattamento al seno	+	34	72	106
	-	207	238	445
		241	310	551



E		M			FREQUENZE ASSOLUTE
		+	-		
+	34	72	106		
-	207	238	445		
		241	310	551	

E		M			STIME DI PROBABILITA'
		+	-		
+	0.06	0.13	0.19	$P(.) = \frac{n(.)}{n}$	
-	0.38	0.43	0.81		
		0.44	0.56	1.00	



Per il calcolo delle probabilità rilevanti si può indifferentemente utilizzare la tabella delle frequenze assolute o quella delle stime di probabilità

esempio: stimate la $P(M+)$ nei primi 4 mesi di vita

		M		
		+	-	
E	+	34	72	106
	-	207	238	445
		241	310	551

		M		
		+	-	
E	+	0.06	0.13	0.19
	-	0.38	0.43	0.81
		0.44	0.56	1.00

$$P(M+) = \frac{241}{551} = 0.44$$

$$P(M+) = P(M+ \cap E+) + P(M+ \cap E-) = 0.06 + 0.38 = 0.44$$



esempio: stimate la $P(M+ \cup E+)$ nei primi 4 mesi di vita

		M		
		+	-	
E	+	34	72	106
	-	207	238	445
		241	310	551

		M		
		+	-	
E	+	0.06	0.13	0.19
	-	0.38	0.43	0.81
		0.44	0.56	1.00

$$P(M+ \cup E+) = \frac{241 + 106 - 34}{551} = 0.57$$

$$P(M+ \cup E+) = P(M+) + P(E+) - P(M+ \cap E+) = 0.44 + 0.19 - 0.06 = 0.57$$





ESERCIZIO:
stimate $P(M+ | E+)$ e $P(M+ | E-)$

		M		
		+	-	
E	+	34	72	106
	-	207	238	445
		241	310	551

		M		
		+	-	
E	+	0.06	0.13	0.19
	-	0.38	0.43	0.81
		0.44	0.56	1.00



SOLUZIONE

- $P(M+ | E+) = P(M+ \cap E+) / P(E+) = (34/551) / (106/551) = 34 / 106 = 0.32$
- $P(M+ | E-) = 207 / 445 = 0.47$



ESERCIZIO:

Nella tabella seguente è riportata la distribuzione di frequenza congiunta del sesso e della capacità vitale forzata (FVC) in cl:

		SESSO		
		Maschi	Femmine	TOTALE
FVC	[200-300]	0	5	5
	(300-400]	4	27	31
	(400-500]	21	13	34
	(500-600]	20	1	21
	(600-750]	9	0	9
TOTALE		54	46	100



Qual è la probabilità che un soggetto abbia un valore dell'FVC > 500 cl?

Qual è la probabilità che un maschio abbia un valore dell'FVC > 500 cl?

Qual è la probabilità che un soggetto abbia un valore dell'FVC > 500 cl e sia femmina?

Qual è la probabilità che un soggetto sia femmina dato che ha un valore dell'FVC ≤ 400 cl?



Qual è la probabilità che un soggetto abbia un valore dell'FVC > 500 cl?

$$P(\text{FVC} > 500) = (21+9) / 100 = 0.30$$

Qual è la probabilità che un maschio abbia un valore dell'FVC > 500 cl?

$$P(\text{FVC} > 500 \mid \text{maschio}) = (20+9) / 54 = 0.54$$

Qual è la probabilità che un soggetto abbia un valore dell'FVC > 500 cl e sia femmina?

$$P(\text{FVC} > 500 \cap \text{femmina}) = (1+0) / 100 = 0.01$$

Qual è la probabilità che un soggetto sia femmina dato che ha un valore dell'FVC ≤ 400 cl?

$$P(\text{femmina} \mid \text{FVC} \leq 400) = (5+27) / (5+31) = 0.89$$



VARIABILE CASUALE:

Si definisce variabile casuale (V.C.) una variabile che assume differenti valori con definite probabilità

ESEMPLI

- Esito del lancio di due dadi

X	2	3	4	5	6	7	8	9	10	11	12
P(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- Numero di aborti su n gravidanze
- Numero di incidenti stradali in un anno in Veneto
- Presenza di asma in un soggetto adulto
- ...



VARIABILE CASUALE DISCRETA:

V.C. che assume un numero discreto (finito o numerabile) di valori

Funzione di probabilità $f(x)$

$$f(x) = \text{Prob}(X = x)$$

dove $f(x)$ è tale che:

1. $0 \leq f(x) \leq 1$ per ogni x

2. $\sum_x f(x) = 1$

- Esempi:
- Numero di volte che compare testa su 100 lanci di una moneta
 - Numero di figli maschi su 4 nascite
 - Numero di decessi in un mese in una popolazione



VARIABILE QUALITATIVA DICOTOMICA

*I fenomeni dicotomici (X) hanno DISTRIBUZIONE BERNOULLIANA
(→ **MODELLO TEORICO**)*

Presenza della caratteristica ($X = 1$):

$$\text{Prob}(X = 1) = \pi$$

Assenza della caratteristica ($X = 0$):

$$\text{Prob}(X = 0) = 1 - \pi$$

**PARAMETRO
(CARATTERISTICA
DELLA POPOLAZIONE):**

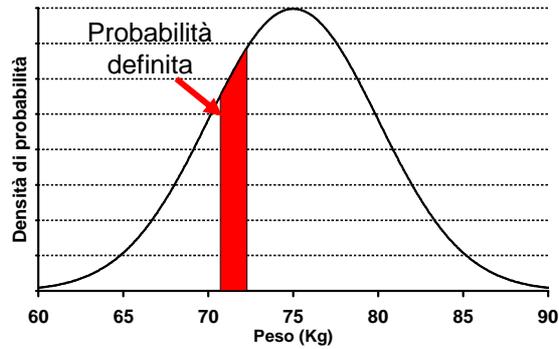
$\pi =$ **probabilità (rischio)
nella popolazione**

Esempio: Presenza di asma in un soggetto adulto (20-44 anni)
residente in Italia nel 1998-2000
→ $p =$ stima di $\pi = 0.049$ [fonte: studio ISAYA]



VARIABILE CASUALE CONTINUA:

V.C. che può assumere infiniti valori all'interno di un intervallo [a, b]



E' definita la probabilità di un intorno (per quanto piccolo) di x:

$$0 \leq \text{Prob}(x-\delta \leq X \leq x+\delta) \leq 1$$



VARIABILE CASUALE CONTINUA:

Funzione di densità di probabilità $f(x)$

dove $f(x)$ è tale che:

1. $f(x) \geq 0$ per ogni x

2. $\int_{-\infty}^{+\infty} f(x) dx = 1$

3. $\text{Prob}(a \leq x \leq b) = \int_a^b f(x) dx$

Esempi:

- Pressione sistolica

- Livello ematico di glucosio

- Altezza

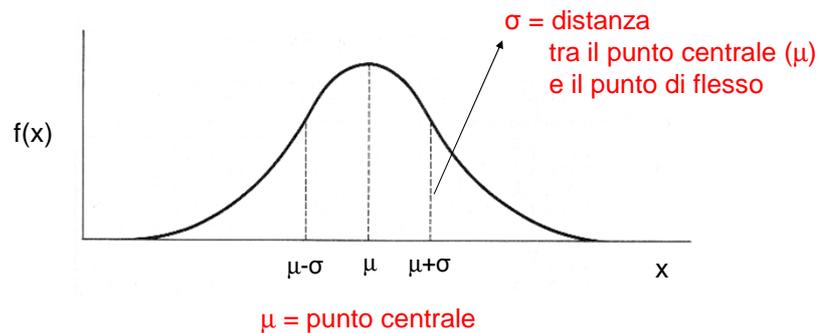
- Peso



VARIABILE QUANTITATIVA (DISTRIBUZIONE SIMMETRICA)

Molte variabili biologiche (X) hanno **DISTRIBUZIONE NORMALE** o approssimativamente normale (\rightarrow **MODELLO TEORICO**)

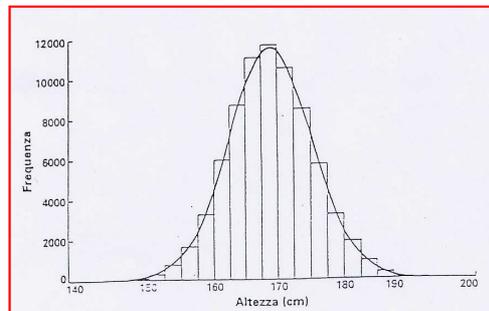
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$



SES

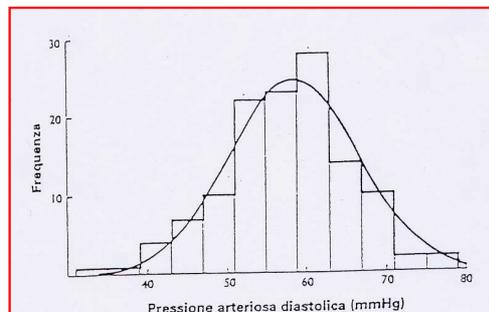
Esempio:

altezza di giovani maschi adulti



Esempio:

pressione arteriosa diastolica del sangue di scolari



SES

