LEZIONI DI STATISTICA MEDICA

Dott. 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*

SESM

SPAZI CAMPIONARI ED EVENTI

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

Esempi:

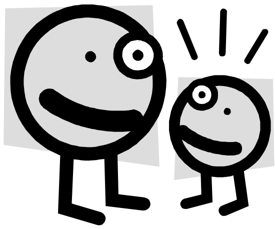
- *estrazione di un numero al lotto*
- *lancio di una moneta*
- *valutazione della presenza di un'infezione virale*

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**.

SESM

Per la rappresentazione degli spazi campionari e dei loro elementi si utilizza la **NOTAZIONE INSIEMISTICA**.



ESPERIMENTO	**SPAZIO CAMPIONARIO**
lancio consecutivo di due monete	**{TT, TC, CT, CC}**
lancio di un dado	**{1, 2, 3, 4, 5, 6}**
misurazione della temperatura corporea	$\{x \mid 34 < x < 42\}$
misurazione del sesso e del fumo in un passante	**{M fum, M non fum, F fum, F non fum}**

E’ possibile definire un qualsiasi evento A come <u>combinazione</u> di più eventi elementari $\rightarrow$ **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}

SESM

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

$$A \subset S$$

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

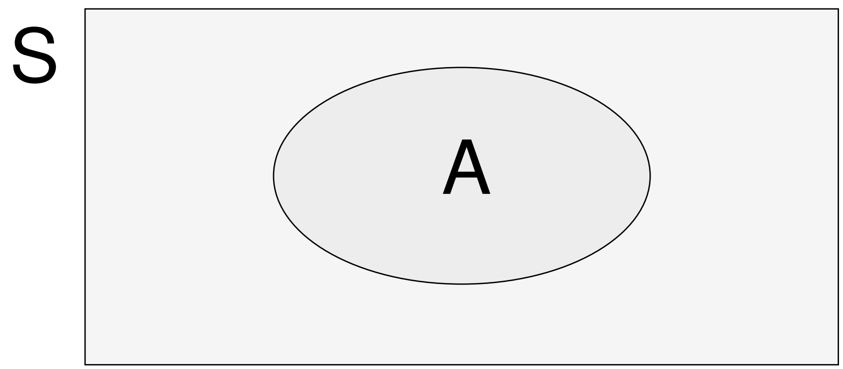
$$\Rightarrow \boxed{A = S}$$

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

$$\Rightarrow \boxed{A = \varnothing}$$

SESM

Gli eventi e gli spazi campionari possono essere rappresentati graficamente mediante i **DIAGRAMMI DI VENN**:



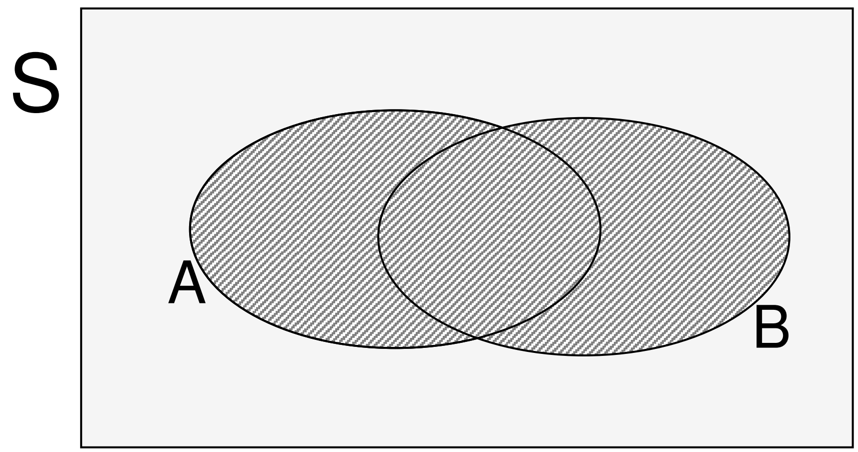
dove l’evento A è il sottoinsieme formato dagli eventi elementari in esso inclusi.

OPERAZIONI LOGICHE SUGLI EVENTI

Dato uno spazio campionario S e degli eventi {A, B, C, ...} 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 **<u>oppure</u>** a B.

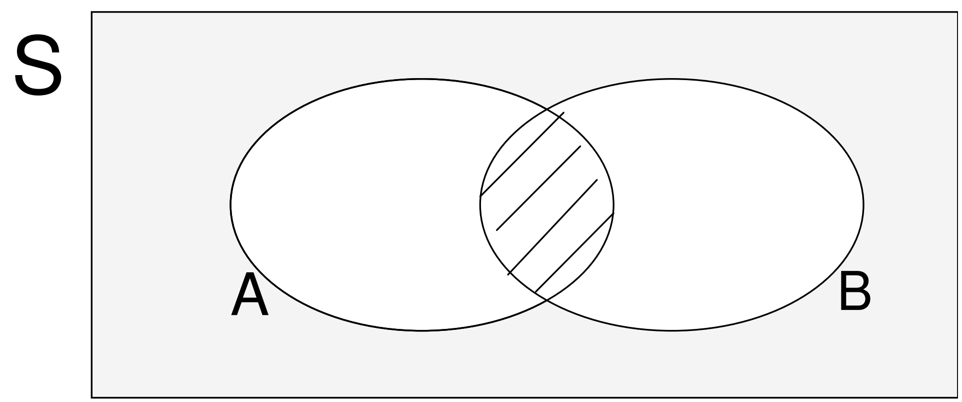


$$C = A \cup B$$

SESM

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.



$$C = A \cap B$$

Esercizio: nel lancio di un dado sia

A = ‘numero pari’
B = ‘numero $\geq$ 4’

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

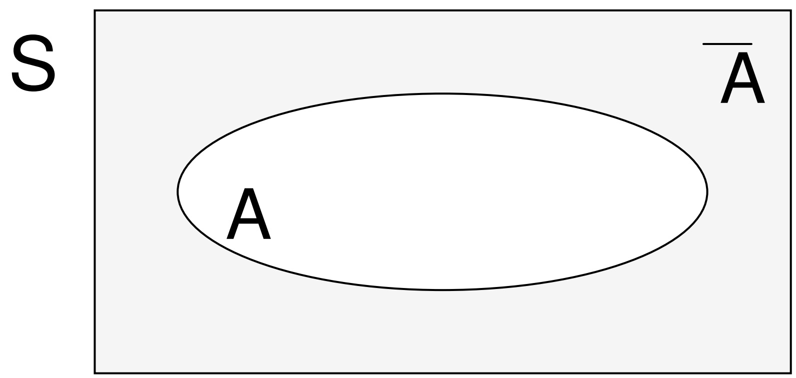
A = {2, 4, 6} **B = {4, 5, 6}**

UNIONE → $A \cup B = \{2, 4, 5, 6\}$

INTERSEZIONE → $A \cap B = \{4, 6\}$

NEGAZIONE DI UN EVENTO: $\overline{A}$

Dato un evento A, la sua negazione identifica un nuovo evento $\overline{A}$ costituito da tutti gli elementi di S non appartenenti ad A. A è detto **complemento di A in S**.



Segue che:

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

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 <u>mutuamente esclusivi</u> $\Rightarrow$ $A \cap B = \varnothing$

PROBABILITÀ

Lo spazio campionario S rappresenta l’insieme dei possibili risultati di un esperimento.

In genere, di fronte alla possibilità di un evento, esprimiamo la nostra maggiore o minore fiducia sul fatto che esso si verifichi

attribuiamo una maggiore o minore probabilità al verificarsi dell’evento

Esempi:

- *è probabile che oggi piova*

- *è più probabile che il tumore al polmone insorga in un fumatore che in un non fumatore*

- *è più probabile che il tumore al seno insorga in una donna che ha partorito per la prima volta dopo i 40 anni*

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)

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

Esempi:

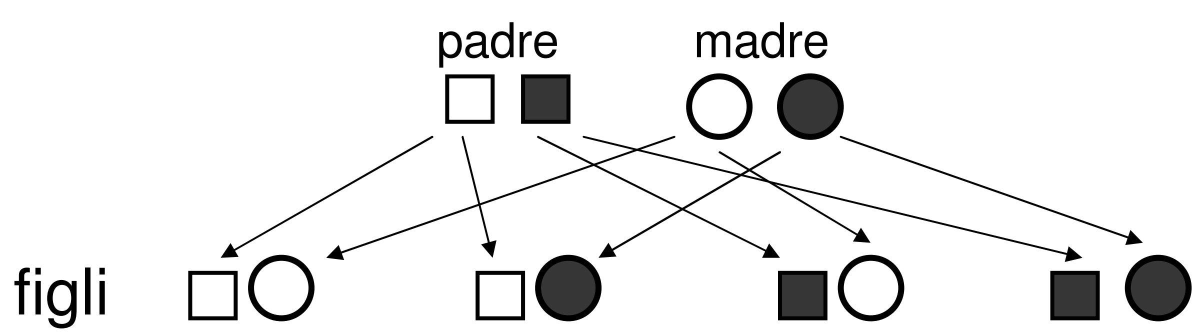
- **probabilità di estrarre un asso da un mazzo di 52 carte = 4/52 = 0.08**
- **probabilità di ottenere testa nel lancio di una moneta = 1/2 = 0.5**

- 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

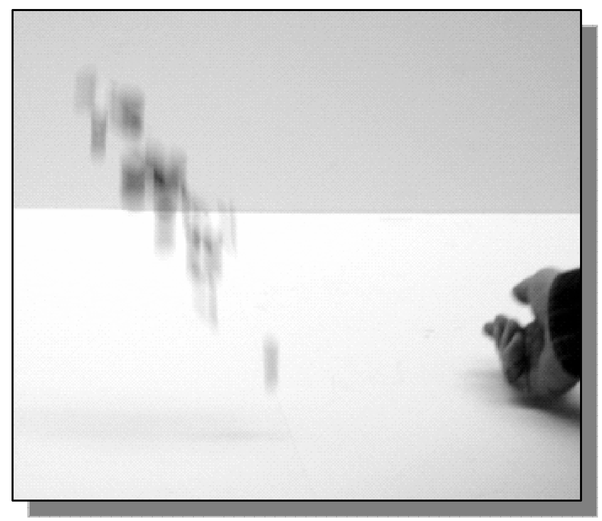
Se entrambi i genitori sono portatori sani del gene della talassemia o della fibrosi cistica, la probabilità di avere un figlio malato è 0.25.



CONCEZIONE FREQUENTISTA DELLA PROBABILITÀ

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

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



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

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

Frequenza relativa

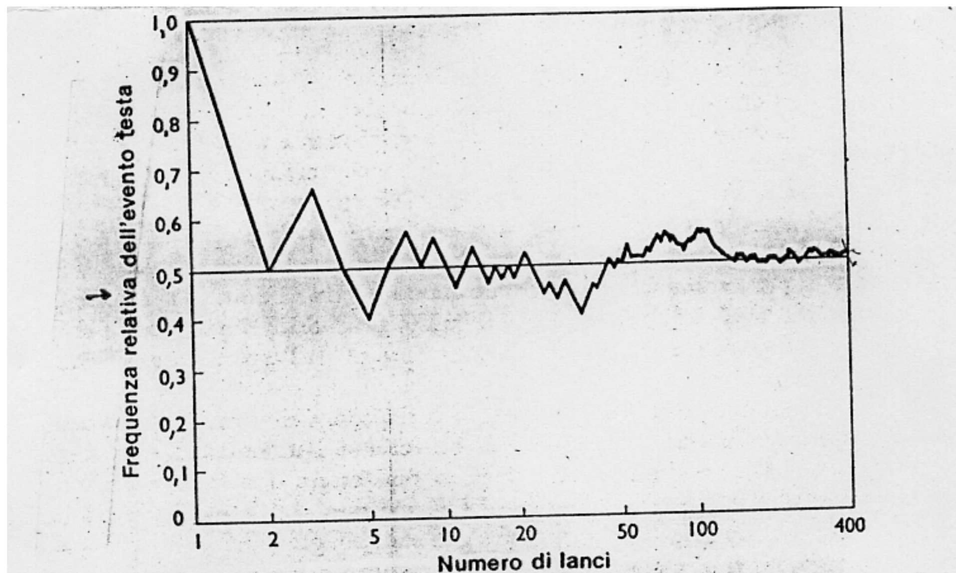


Fig. 2.1. Frequenza dell'evento testa in una successione di lanci di una moneta; scala logaritmica per il numero di lanci (da Cramer, 1946, per concessione dell'autore e degli editori).

Numero di lanci

Esempio 2:

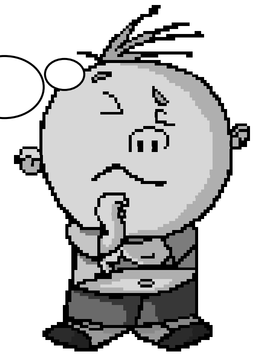
Prob. che un bambino italiano nasca morto nel 2000 $= \dfrac{\textit{numero nati morti nel 2000}}{\textit{numero nati nel 2000}}$

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

gravidanza

maschio **femmina**

1 su 2 = 50%



(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 <u>funzione a valori reali</u> definita sullo spazio campionario S che soddisfa i seguenti assiomi:

1) per qualsiasi evento $A \subset S$, si ha che: $\boxed{0 \leq P(A) \leq 1}$

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

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

2) $\boxed{P(S) = 1}$

SESM

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

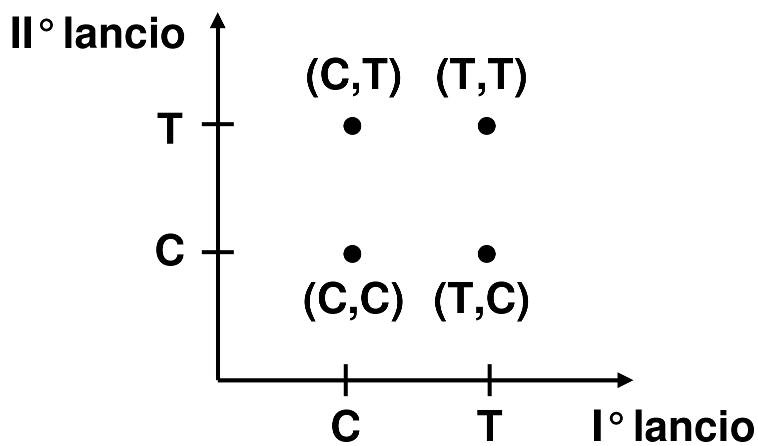
$$\boxed{\mathbf{P(A_1 \cup A_2 \cup \ldots \cup A_i \cup \ldots) = P(A_1) + P(A_2) + \ldots + P(A_i) + \ldots}}$$

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

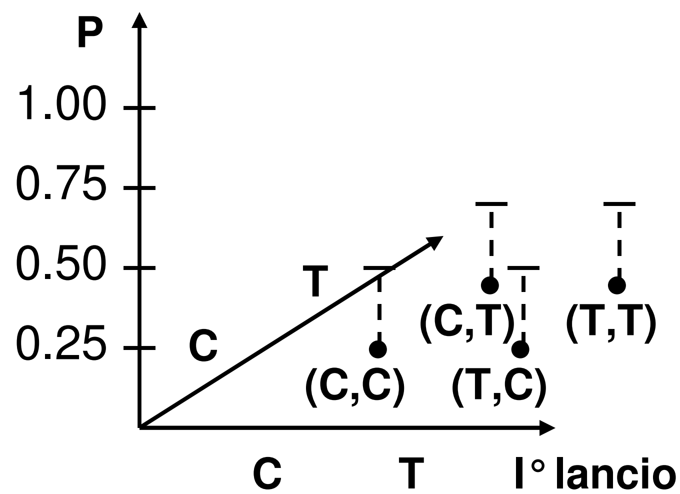
$\Rightarrow$ la probabilità corrispondente ad un qualsiasi **evento composto A** sarà definita come <u>somma delle probabilità degli eventi elementari contenuti in A</u> (*assioma 3*).

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'



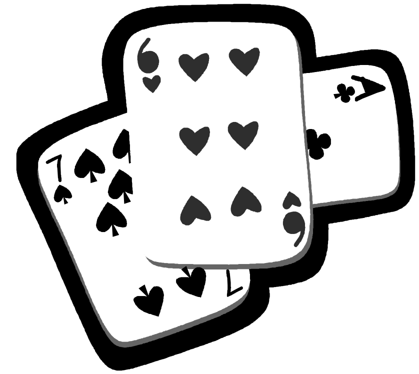
Probabilità di ottenere almeno una testa

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

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

MAZZO DI 52 CARTE

ESERCIZIO:



1. calcolare la probabilità di estrarre <u>l'asso di cuori</u>

2. calcolare la probabilità di estrarre <u>una carta rossa</u>

3. calcolare la probabilità di estrarre <u>una figura</u>

4. calcolare la probabilità di estrarre <u>un re</u>

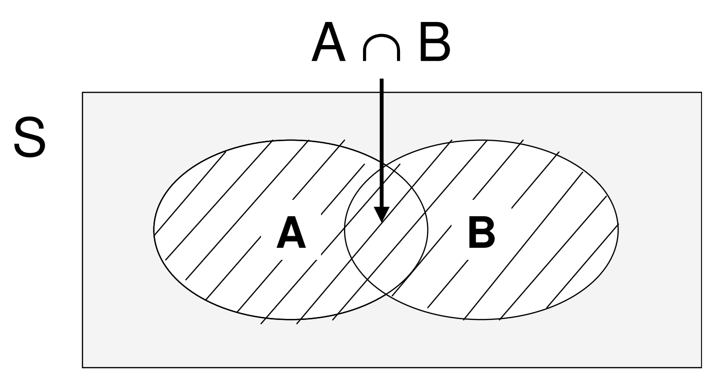
REGOLE DEL CALCOLO DELLA PROBABILITÀ

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**):

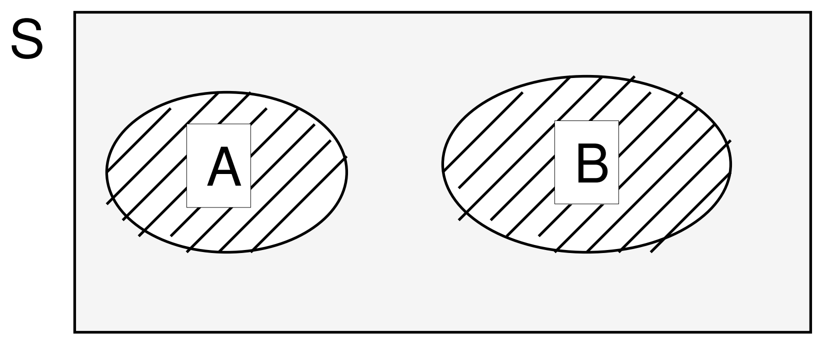
$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$



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)

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

Se Ā è il complemento di A in S:

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

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

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

SESM

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

→ **PROBABILITA' CONDIZIONALE**

Esempio:

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

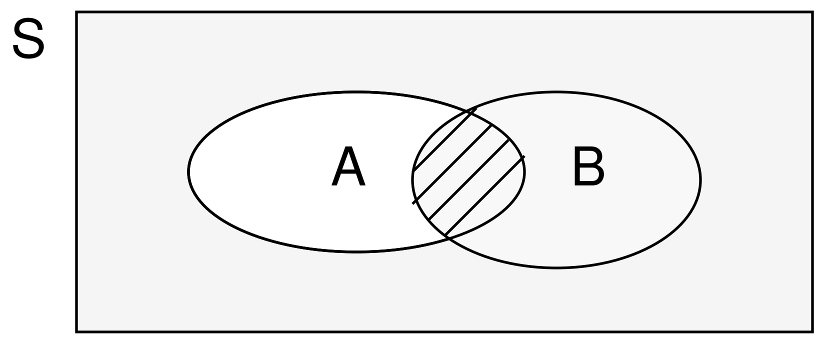
probabilità di avere un tumore al polmone dato che si fuma

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

SESM

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

$$\boxed{P(A \mid B) = P(A \cap B) \,/\, P(B)}$$



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

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

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

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

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

$$P(A \mid B) = P(A)$$

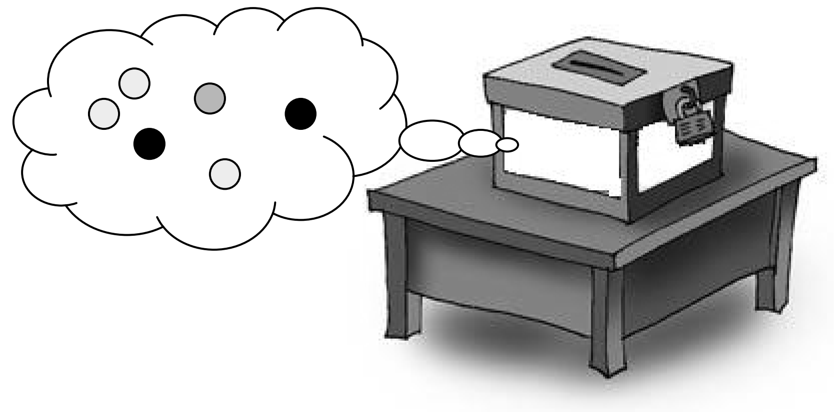
e i due **eventi** sono detti **indipendenti**, ovvero:

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

SESM

Esempio:

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



A = estraz. I^a pallina gialla → $P(A) = 3/6$

B = estraz. II^a pallina gialla → $P(B|A) = 2/5$

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

Se l'estrazione fosse <u>con reimbussolamento</u>:

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

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

CALCOLO DEGLI ATTESI

Definita la probabilità di un evento o di una qualsiasi combinazione di eventi, è immediato definire il **numero di eventi attesi** in una serie di prove ripetute in modo casuale.

DEFINIZIONE: Se P(A) è la probabilità di comparsa dell'evento A, il **numero di eventi attesi E** in una serie di N prove sarà dato da:

$$E = P(A) \cdot N$$

Esempio: *Se la probabilità di influenza per gli adulti tra i 20 e i 40 anni in un determinato mese dell'anno è* ***0.20****, quanti malati ci attendiamo in un campione casuale di* ***800*** *persone?*

$$\mathbf{E = 0.20 \cdot 800 = 160}$$

(numero atteso di soggetti con l'influenza)

Esempio: *Si supponga che la probabilità di fumare sia **0.30** e quella di avere un tumore al polmone sia **0.01**. In un'indagine condotta su **1000** individui, sono stati osservati **15** soggetti fumatori con tumore al polmone.*

Secondo voi, i dati dimostrano che avere il tumore al polmone e fumare sono eventi indipendenti?

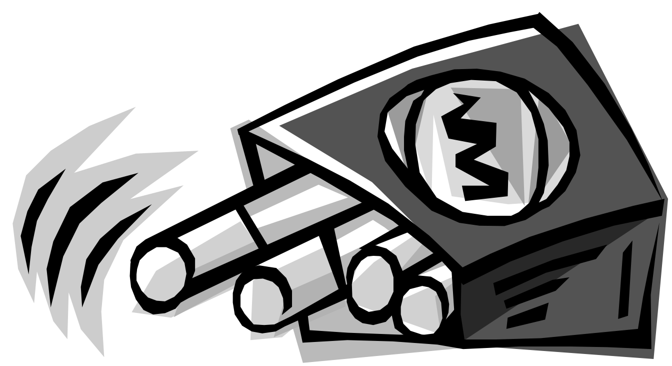
E = (0.30 • 0.01) • 1000 = 3

(numero atteso di fumatori malati sotto l'ipotesi di indipendenza)

O = 15

(numero osservato di malati fumatori)

$\rightarrow$ **eccesso di 12 casi**



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):

- $P(M+ \mid E+)$
- $P(M+ \mid E-)$

SESM

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

SESM

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:**

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

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

FREQUENZE ASSOLUTE

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

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

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

STIME DI PROBABILITA'

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

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

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

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

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

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

SESM

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

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

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

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

$$P(M+ \cup E+) = P(M+) + P(E+) - P(M+ \cap E+) = \mathbf{0.44} + \mathbf{0.19} - \mathbf{0.06}$$

$$= 0.57$$

SESM

ESERCIZIO:

stimate P(M+ | E+) e P(M+ | E-)

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

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

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 $\leq$ 400 cl?

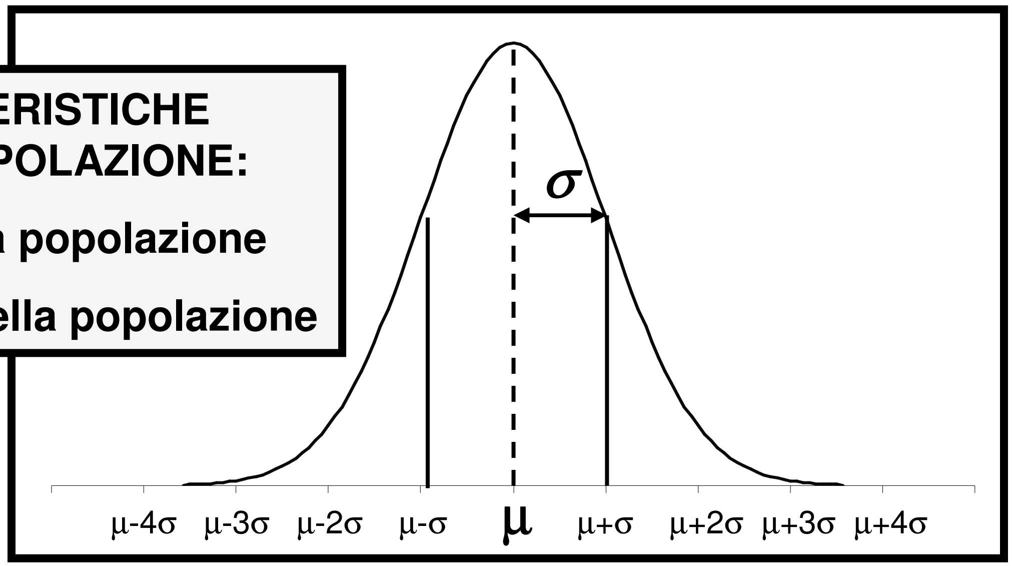
OUTCOME QUANTITATIVO (DISTRIBUZIONE SIMMETRICA)

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

CARATTERISTICHE DELLA POPOLAZIONE:

μ **= media nella popolazione**

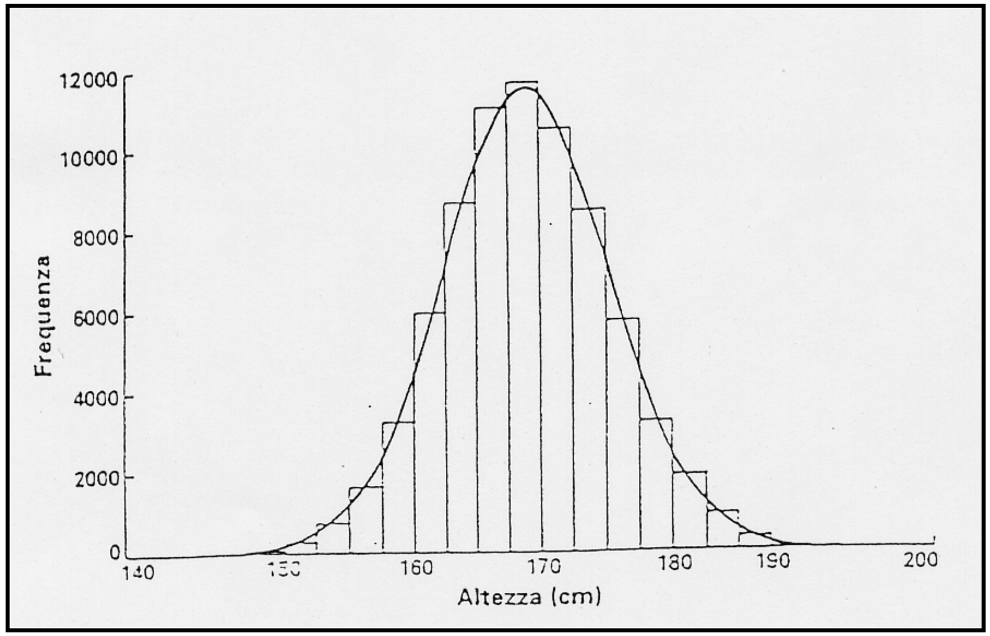
σ **= dev. std. nella popolazione**



Prob($\mu-\sigma < X < \mu+\sigma$) = **0.68**

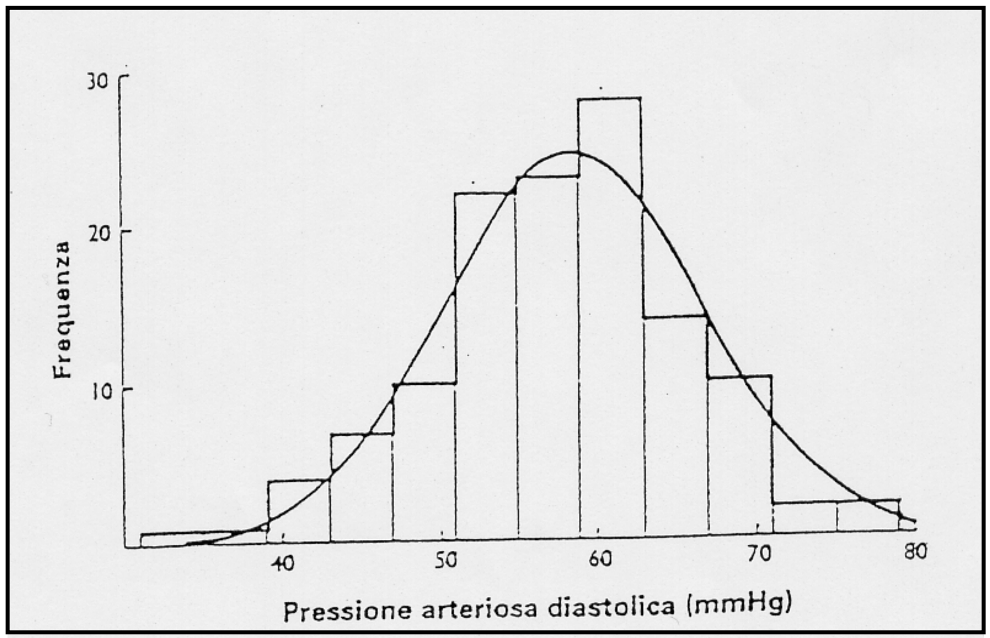
Esempio:

altezza di giovani maschi adulti

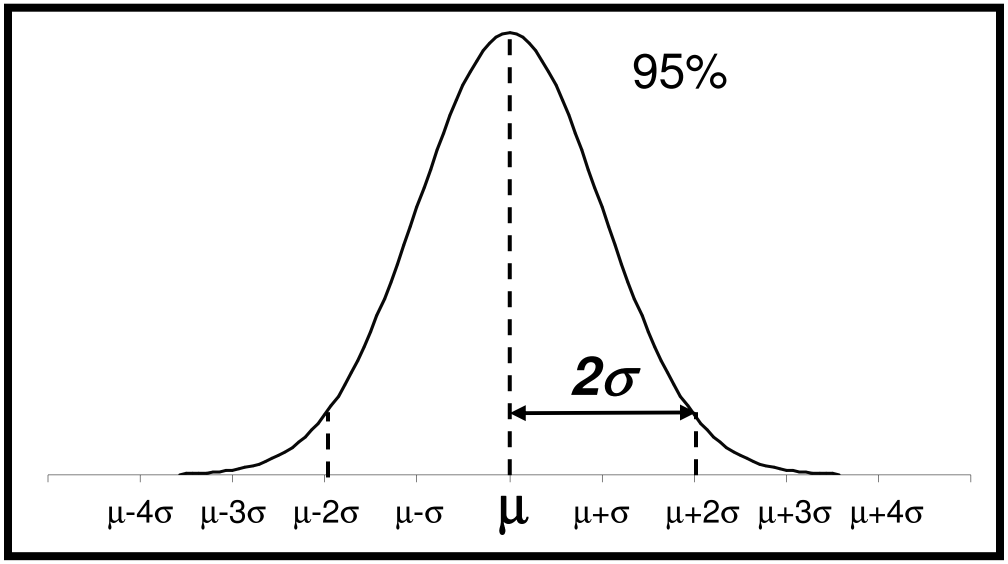


Esempio:

pressione arteriosa diastolica del sangue di scolari



VALORI DI NORMALITA'



2.89 **7.84** **mmol/L**

Esempio:

I valori di normalità dell'UREA sono compresi tra **2.89** e **7.84** mmol/L.