

Metodi statistici per lo studio dei fenomeni biologici

Alla fine di questa lezione dovrete essere in grado di:

- ❑ spiegare il concetto di probabilità
- ❑ spiegare il concetto di probabilità condizionata
- ❑ spiegare il significato di eventi indipendenti e non indipendenti
- ❑ eseguire le operazioni sulla probabilità



Probabilità

Nel foglietto illustrativo di un farmaco antinfiammatorio, è riportata la seguente dicitura:

ATTENZIONE: C'è una probabilità del 15% di avere una rash cutaneo (irritazione della pelle). In caso di arrossamento della pelle consultare il medico.

Come interpretare al meglio questo avvertimento?

- a. Evitare di usare il farmaco sulla pelle, c'è una discreta probabilità di sviluppare un rash cutaneo
- b. Se usato sulla pelle, va applicato solo il 15% della dose raccomandata
- c. Se si presenta un rash cutaneo, probabilmente interesserà solo il 15% della pelle su cui il farmaco è applicato
- d. Circa 15 persone su 100 che usano il trattamento presenteranno un rash cutaneo
- e. Usando questo trattamento, difficilmente si avrà un rash cutaneo

Probabilità: Definizioni

classica

La probabilità di un evento è il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili, purché questi ultimi siano tutti ugualmente probabili.

frequentista

La probabilità di un evento è il limite della frequenza relativa dei successi, quando il numero delle prove tende all'infinito

soggettiva

La probabilità è il grado di fiducia che una persona ha nel verificarsi di un evento

Probabilità: Definizioni

classica

La probabilità di un evento è il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili, purché questi ultimi siano tutti ugualmente probabili.

frequentista

La probabilità di un evento è il limite della frequenza relativa dei successi, quando il numero delle prove tende all'infinito

soggettiva

La probabilità è il grado di fiducia che una persona ha nel verificarsi di un evento

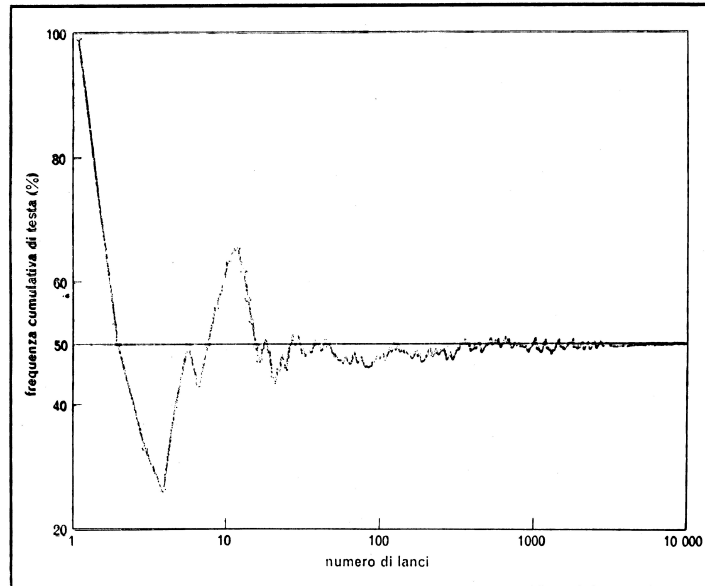


Fig.1 Frequenza della risposta 'testa' aumentando il numero di lanci. Aumentando il numero delle osservazioni, la frequenza relativa tende a un valore limite stabile, valore approssimato della probabilità di ottenere la risposta 'testa' in un singolo lancio della moneta.

Probabilità

Quale delle seguenti sequenze è più probabile lanciando una moneta 5 volte?

- a. TTTCC
- b. CTTCT
- c. CTCCC
- d. TCTCT
- e. Le quattro sequenze hanno tutte la stessa probabilità

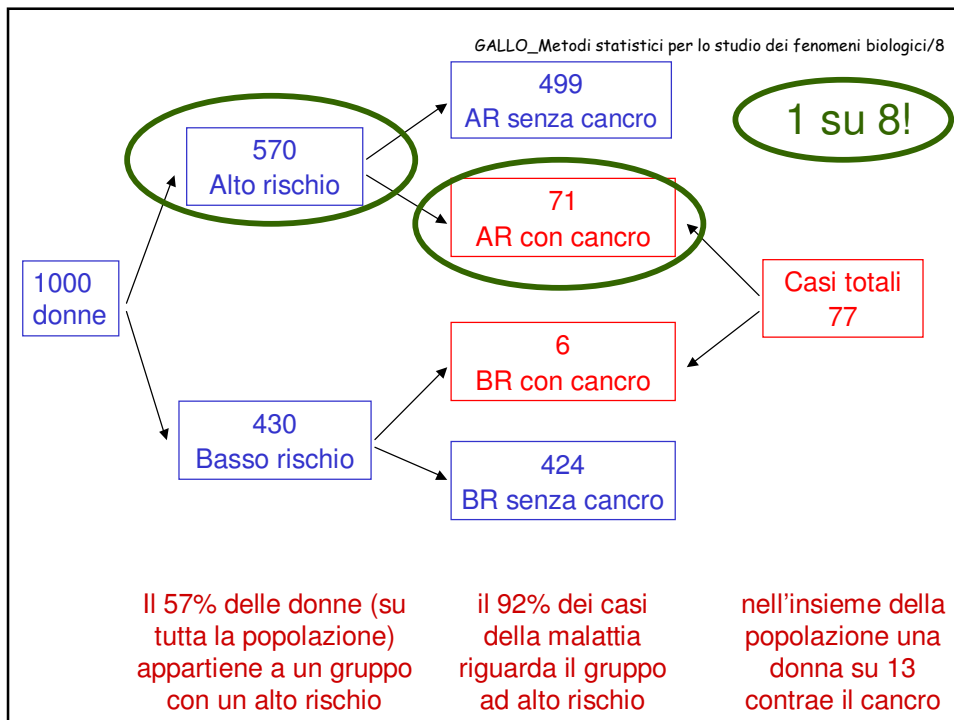
Probabilità

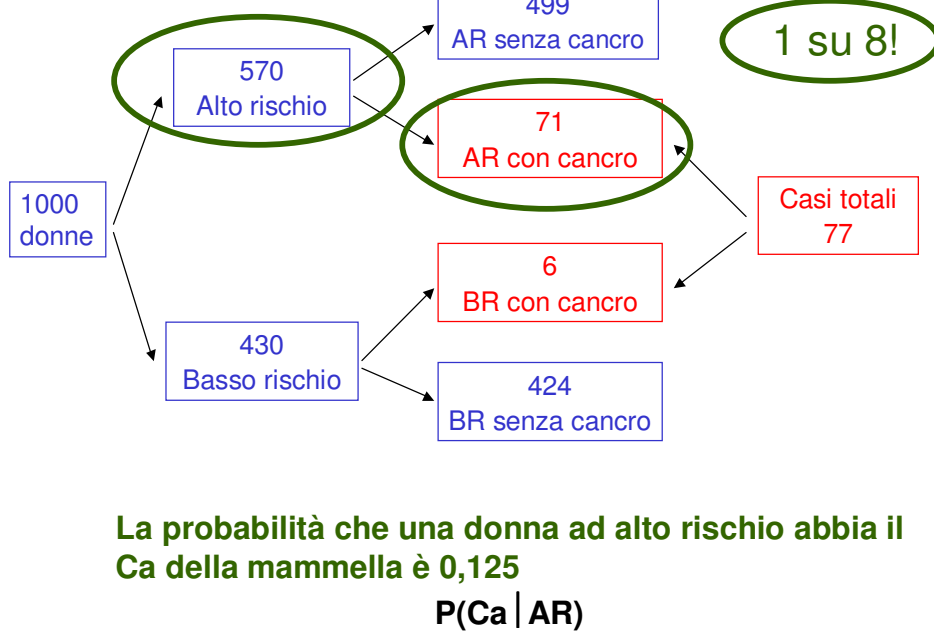
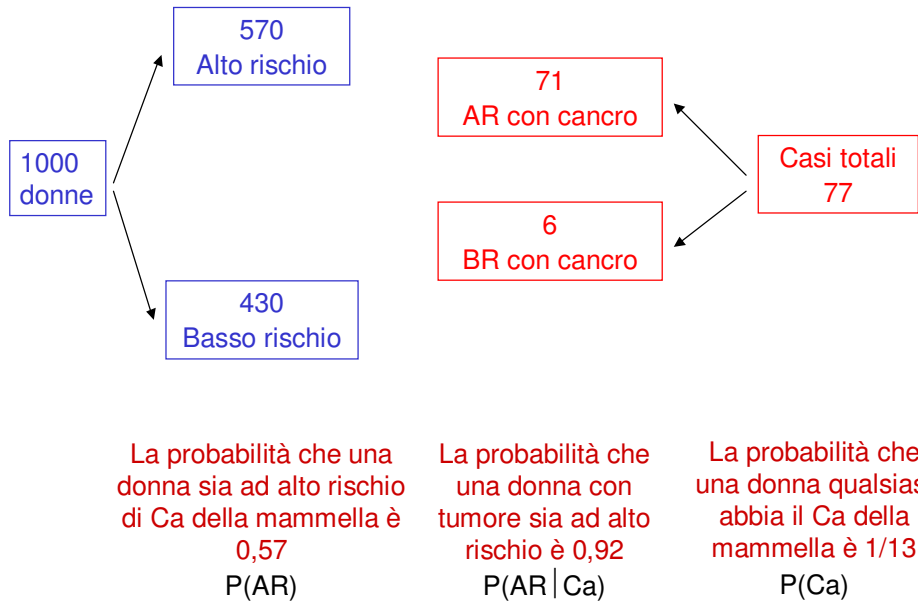
Un chirurgo del Michigan consigliava la mastectomia profilattica (cioè l'asportazione del seno in donne sane) nelle donne con più di 30 anni di età, ad alto rischio di contrarre il cancro della mammella. Questa affermazione si basava sul seguente ragionamento:

“Il 57% delle donne (su tutta la popolazione) appartiene a un gruppo con un alto rischio di contrarre il cancro della mammella, tanto che il 92% di tutti i casi della malattia riguardano proprio questo gruppo. Inoltre, nell'insieme della popolazione (ad alto e basso rischio) una donna su 13 contrae il cancro al seno fra i 40 e i 59 anni. Se ne conclude che, nel gruppo ad alto rischio, fra i 40 e i 59 anni, si ammala di cancro al seno una donna ogni 2 o 3”.

Ammettendo che questi dati siano giusti (in realtà, una stima migliore nelle donne fra 40 e 59 anni è 1/36), il ragionamento è corretto?

Gigerenzer 2002



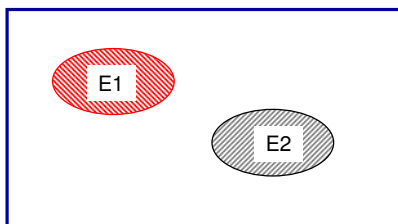


Probabilità

- La probabilità che si verifichi almeno uno dei risultati possibili è uguale a 1
- La probabilità che non si verifichi nessuno dei risultati possibili è uguale a 0
- La probabilità che si verifichi un particolare risultato fra quelli possibili è compresa fra 0 e 1
- La probabilità che non si verifichi un particolare risultato è uguale a 1 meno la probabilità che si verifichi

Operazioni sulla probabilità

La somma

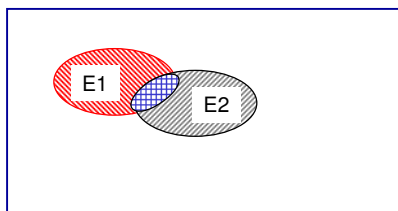


$$P(E_1 \cup E_2) = P(E_1 \circ E_2) =$$

eventi che si escludono

$$= P(E_1) + P(E_2)$$

$$P(AR) + P(BR) = 0,57 + 0,43 = 1$$



eventi che **non** si escludono

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

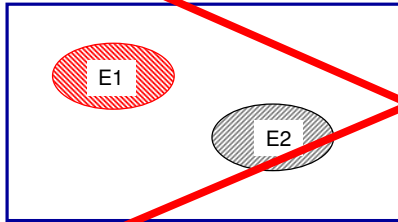
$$P(AR) + P(Ca) =$$

$$0,57 + 0,077 - 0,071 = 0,576$$

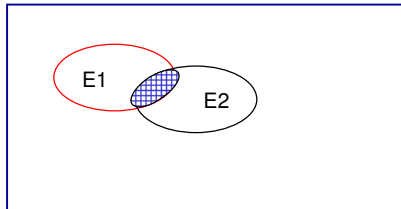
Operazioni sulla probabilità

Il prodotto

$$P(E_1 \cap E_2) = P(E_1 \text{ e } E_2) =$$



eventi che si escludono



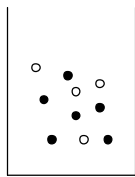
eventi che **non** si escludono

Qual è la probabilità dell'evento congiunto?

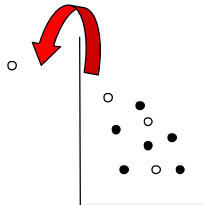
Eventi congiunti

E_1 = estrarre una pallina bianca

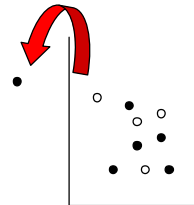
E_2 = estrarre una pallina bianca alla seconda estrazione



$$P(E_1) = 4/10$$



$$P(E_2 | E_1) = 3/9$$



$$P(E_2 | E_1) = 4/9$$

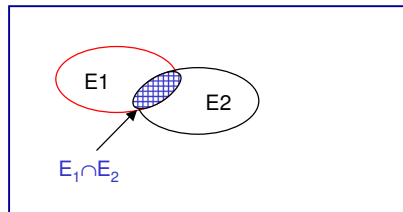
La probabilità del secondo evento varia in funzione del verificarsi o meno del primo → **probabilità condizionata**

$P(E_2 | E_1) \neq P(E_2)$ → **eventi associati (o non indipendenti)**

$P(E_2 | E_1) = P(E_2)$ → **eventi indipendenti**

Operazioni sulla probabilità

Il prodotto



$$P(E_1 \cap E_2) = P(E_1 \text{ e } E_2) =$$

Eventi non indipendenti

$$= P(E_1) * P(E_2 | E_1)$$

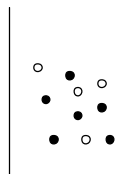
$$P(\text{Ca e AR}) = P(\text{Ca}) * P(\text{AR} | \text{Ca}) \\ = 0,077 * 0,92 = 0,071$$

$$= P(E_2) * P(E_1 | E_2)$$

$$P(\text{Ca e AR}) = P(\text{AR}) * P(\text{Ca} | \text{AR}) \\ = 0,57 * 0,125 = 0,071$$

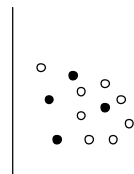
Eventi congiunti

E_1 = estrarre una pallina
bianca



$$P(E_1) = 4/10$$

E_2 = estrarre una pallina
bianca

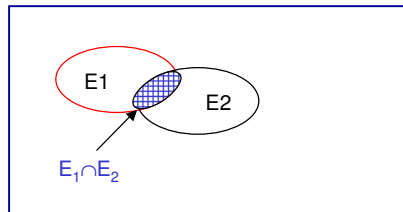


$$P(E_2) = 8/12$$

Il verificarsi di un evento **non** modifica la probabilità del verificarsi dell'altro → **Eventi indipendenti**

Operazioni sulla probabilità

Il prodotto



$$P(E_1 \cap E_2) = P(E_1 \text{ e } E_2) =$$

Eventi indipendenti

$$= P(E_1) * P(E_2)$$

poiché $P(E_2 | E_1) = P(E_2)$

Non possiamo sapere 'a priori' se due eventi sono indipendenti, se non per disegno

All'inverso, possiamo utilizzare la stima in condizioni di indipendenza come 'riferimento, per valutare se due eventi sono indipendenti

Probabilità e diagnosi

Ca mammario			
Rischio	SI	NO	Totale
Alto	71	499	570
Basso	6	424	430
Totale	77	923	1000

Qual è la probabilità che una donna sia ad alto rischio e abbia un cancro?


$$P(\text{Ca e AR}) = 0,071 = P(\text{Ca}) * P(\text{AR} | \text{Ca}) = (77/1000) * (71/77)$$

Qual è la probabilità che una donna ad alto rischio abbia un cancro?

$$P(\text{Ca} | \text{AR}) = 0,125 = (71/1000) / (570/1000) = 71/570$$

Probabilità

Quale delle seguenti sequenze è più probabile lanciando una moneta 5 volte?

- a. TTTCC $= 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5$
- b. CTTCT $= 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5$
- c. CTCCC $= 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5$
- d. TCTCT $= 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5$
-  e. Le quattro sequenze hanno tutte la stessa probabilità

Studenti I anno 2006-2007

Maschi			Femmine		
gruppo	n	%	gruppo	n	%
O	36	46.8	O	47	52.2
A	20	26.0	A	20	22.2
B	11	14.3	B	15	16.7
AB	10	13.0	AB	8	8.9
totale	77	54.6	totale	90	70.3
mancante	64	45.4	mancante	38	29.7

Studenti I anno 2006-2007

Gruppo	Sesso		Totale
	Maschile	Femminile	
O	36	47	83
A	20	20	40
B	11	15	26
AB	10	8	18
Totale	77	90	167

Qual è la probabilità che un soggetto sia di sesso maschile e di gruppo O ?

$$P(M \cap O) = 36/167 = (83/167) * (36/83) = (77/167) * (36/77)$$

Studenti I anno 2006-2007

Gruppo	Sesso		Totale
	Maschile	Femminile	
O	36	47	83
A	20	20	40
B	11	15	26
AB	10	8	18
Totale	77	90	167

Se sesso e gruppo sanguigno fossero indipendenti, quanti studenti dovremmo vedere di sesso maschile e di gruppo O ?

$$P(M \cap O) = (83/167) * (77/167) * 167 = 38,3 \neq 36$$