Lezione V:

Probabilità

Cattedra di Biostatistica – Dipartimento di Scienze Biomediche, Università degli Studi "G. d'Annunzio" di Chieti – Pescara

Prof. Enzo Ballone

Lezione 4a- Calcolo delle probabilità

Introduzione.

- Il CdP trova molte applicazioni in Medicina, in Biologia, nelle Scienze Sociali, etc.
- In Genetica, ad esempio, mediante il CdP si possono formulare in maniera più appropriata le leggi di Mendel sull'ereditarietà, sulla ricorrenza di mutazioni, etc;
- In Medicina si può valutare l'efficacia diagnostica di un test, di un esame di laboratorio, del rischio di malattie, di efficacia di trattamenti, etc, attraverso il CdP.

Introduzione.

- L’attività del medico si basa molto sul CdP, specialmente nella parte diagnostica e nella prognosi delle malattie.
- I test diagnostici raramente sono perfettamente accurati, tuttavia quando essi risultano positivi tendono ad aumentare la probabilità nell’individuazione della la malattia; in certo qual modo il medico, attraverso la conoscenza di certe probabilità, ritiene che la malattia possa essere presente o meno.

Introduzione.

- La probabilità di malattia dà una misura numerica di quanto il medico crede che essa sia presente: la malattia è nel paziente, la probabilità nella nostra mente.
- ESPERIMENTO O PROVA:
- è una operazione il cui risultato non è prevedibile con certezza;
- è una osservazione che dà luogo al verificarsi di “fatti aleatori”.

Introduzione.

- Il CdP permette di descrivere e studiare gli eventi aleatori.
- Un evento è aleatorio quando non è possibile prevedere con certezza se si avvererà o meno
- EVENTO: ogni possibile risultato di un esperimento, di un avvenimento; in sostanza un fatto che, a seguito di una prova, può accadere oppure no (incertezza).

Introduzione.

- Quindi, ad ogni prova sono associati degli eventi, ad esempio:
- alla prova “conclusione di una gravidanza” si associano gli eventi “aborto”(A), “nato morto”(NM),“nato vivo”(NV);
- alla prova “numero di nati in unico parto” si associano gli eventi “1”, “2”, “3”;

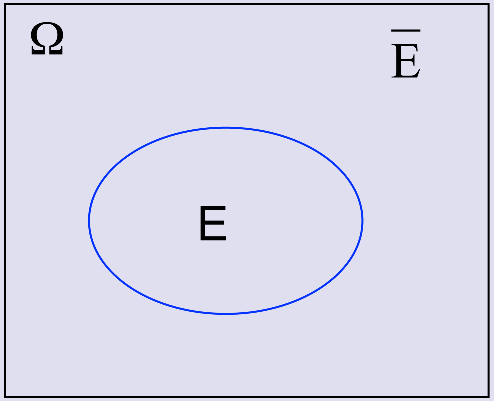
Introduzione.

- alla prova “sesso di un neonato” si associano gli eventi “maschio”(M) o “femmina”(F);
- alla prova “dimissione di un paziente da un ospedale dopo un ricovero” si associano gli eventi “guarito”(G), “non guarito”(NG), “deceduto”(D).
- Tali esiti possono costituire gli eventi elementari dell’esperimento (o prova) preso in esame, e da essi si origina lo **spazio degli eventi** (Ω).

Insiemi - Eventi – Operazioni.

- Un Insieme è un raggruppamento di individui, oggetti o altro che soddisfano una data legge di “associazione”, “appartenenza”

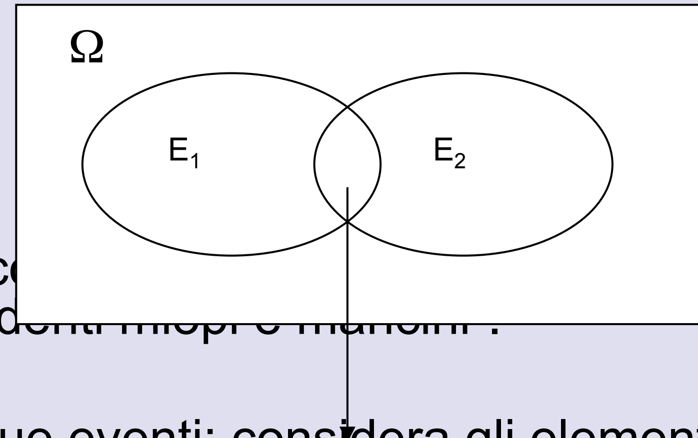
- Ω:Tutto il personale sanitario;
- E: personale medico
- $\overline{E}$: personale paramedico
- $\overline{E}$ = (Ω - E), complementare dell’insieme E
- (E o $\overline{E}$) = unione di E e $\overline{E}$ = Ω



- Gli eventi E possono essere rappresentati con insiemi contenuti in un altro insieme Ω che sta ad indicare lo spazio degli eventi ed è individuabile come l’evento certo; alle operazioni tra insiemi si possono far corrispondere quelle tra eventi.

Intersezione, Unione e Differenza tra insiemi.

- Ω: Studenti di Medicina
- E1: Studente miope
- E2: Studente mancino



- **<u>Intersezione</u>:** E1 **<u>e</u>** E2 (prodotto logic… o gli elementi comuni; nell'esempio "studenti miopi e mancini".

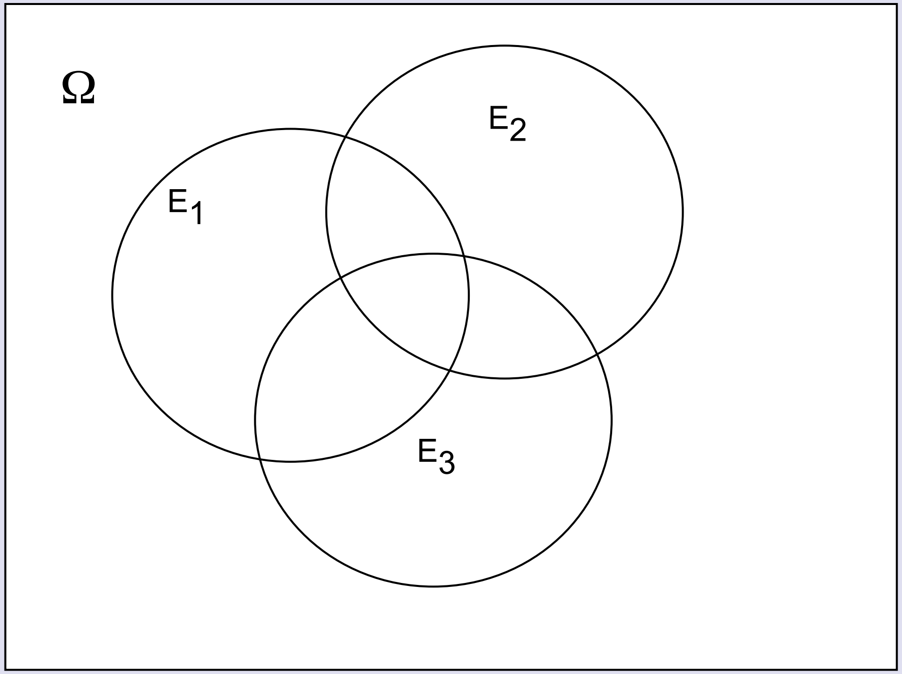
- **<u>Unione</u>** E1 **<u>o</u>** E2 (somma logica) tra due eventi: considera gli elementi di entrambi gli insiemi; nell'esempio "studenti miopi o (or) mancini".

- **<u>Complementare</u>** dell'insieme E1 o **contrario** (negazione "non") dell'evento E1=(Ω- E1); nell'esempio "studenti non miopi".

- **<u>Differenza</u>** E1 - E2 considera gli elementi di uno dei due insiemi escludendo quelli dell'altro; nell'esempio, "miopi, ma non mancini"; mentre l'insieme E2 – E1 indica "studenti mancini, ma non miopi".

Esempio:

- Ω : Tutti i pazienti diabetici (pz)
- E1: Pazienti con Retinopatia Diabetica (RD)
- E2: Pazienti con Nefropatia Diabetica (ND)
- E3: Pazienti con Vasculopatia Cerebrale (VC)



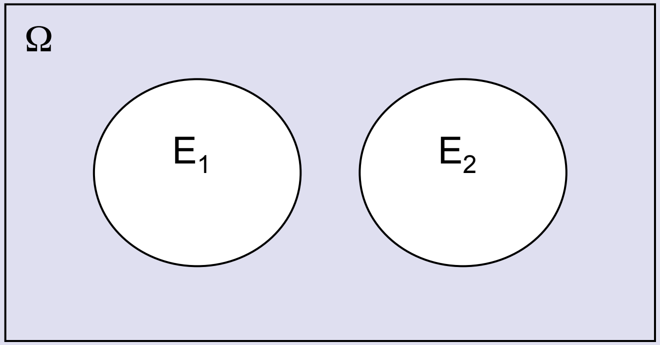
- (E1 e E2 e E3) = insieme dei pz con RD, ND e VC;
- (E1 o E2 o E3) = insieme dei pz che hanno <u>almeno una</u> delle tre complicanze;
- Ω - (E1 o E2 o E3) = insieme dei pazienti diabetici <u>senza complicanze.</u>

Eventi Semplici ed eventi complessi.

- Ad ogni prova è associato un "*universo delle possibilità*", l'insieme dei casi possibili, indicato con
- Ω = ***spazio degli eventi***. Ad esempio:
- alla prova "conclusione di una gravidanza" si associa $\Omega = \{A, NM, NV\}$;
- alla prova "sesso di un neonato" $\Omega = \{M,F\}$;
- alla prova "numero di nati in unico parto" si associa l'insieme $\Omega = \{1,2,3\}$;
- Ognuno dei casi possibili presi in esame rappresenta un ***evento semplice***. Questi possono essere composti mediante i connettivi "e", "o", "non", " - " generando in tal modo ***eventi complessi***, ad esempio: (E1 o E2), (E1 e E2) ; (E1 - E2); (E2 – E1).

Eventi compatibili ed eventi incompatibili.

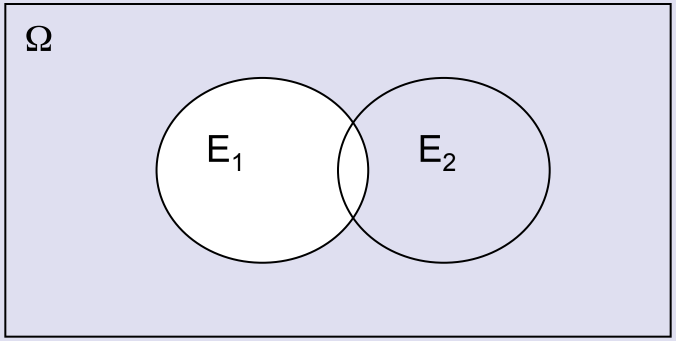
- Due eventi E1 e E2 si dicono ***incompatibili*** se il verificarsi dell'uno esclude il verificarsi dell'altro, cioè il prodotto logico dei due eventi è l'***evento impossibile*** Ø: (E1 e E2) = Ø



- Ad es. nella prova "<u>sesso di un neonato</u>" gli eventi E1 = M ed E2 = F sono incompatibili;
- nella prova "<u>conclusione di una gravidanza</u>" gli eventi E1 = A, E2 = NM, E3 = NV sono incompatibili.

Eventi compatibili ed eventi incompatibili.

- Mentre gli eventi complessi (F e NV), (M e NM) sono eventi compatibili, in quanto possono verificarsi contemporaneamente nella stessa prova.
- In generale due eventi E1 ed E2 sono compatibili se (E1 e E2) ≠ Ø



- Più eventi sono necessari, quando in ogni prova uno di essi deve necessariamente verificarsi
- Più eventi sono incompatibili e necessari, quando in ogni prova uno ed uno solo tra essi deve verificarsi.

La Misura della Probabilità.

- Per probabilità di un evento si intende quel numero che esprime una misura della “possibilità” del verificarsi dell’evento stesso. In sostanza è una misura del “grado di fiducia” sul verificarsi dell’evento.
- Esistono diversi modi di impostare la teoria della misura di probabilità; ognuno di essi porta ad una determinata “definizione” di probabilità. Si farà riferimento alle definizioni: classica, frequentistica e soggettivista.

1-Definizione Classica.

- **Definizione classica (la Prob. dei giochi):** la probabilità di un evento E, si scrive P(E), è data dal rapporto fra il numero dei casi favorevoli al verificarsi di E (K) ed il numero dei casi possibili (n), purché questi ultimi siano tutti egualmente probabili **(EQUIPROBABILI)** (Pari opportunità!), (non sono equiprobabili gli esiti possibili di una partita di calcio)
- $P(E) = \frac{n^\circ \text{ casi favorevoli}}{n^\circ \text{ casi possibili}} = \frac{K}{n}$ con $K \leq n$
- Se k**=0** nessun caso favorevole, allora **P(E)=0**
- Se **k=n** tutti i casi sono favorevoli, allora **P(E)=1 e l'evento è certo**
- Dunque: $\mathbf{0 \leq P(E) \leq 1}$
- Questa definizione si applica solo se: **Ω sia finito e gli eventi siano equiprobabili**. Ciò la rende poca adatta per la maggior parte delle applicazioni; si preferisce, quindi, adottare un diverso punto di vista, ossia:

2-Definizione frequentistica.

- **Definizione frequentistica:** la probabilità di un evento E è stimata attraverso la frequenza relativa di successi (verificarsi di E) osservata in un ***numero"grande"*** di prove o osservazioni, ripetute sotto identiche condizioni.
- E' il limite del rapporto tra il numero r di esiti favorevoli ad E ed il numero totale **n** di prove effettuate, quando il numero di prove diverge:

$$P(E) = \lim_{n \to \infty} \frac{r}{n}$$

- dove il limite va inteso non nel senso della matematica, ma esprime semplicemente il fatto che la frequenza relativa r/n su un gran numero di prove fornisce una "buona" stima della probabilità (legge empirica del caso).

2-Definizione frequentistica.

- N.B.: Mentre nella Def. Classica la Prob. Viene valutata o stabilita a PRIORI, nella frequentista viene ricavata a posteriori, dall'esame dei dati statistici.
- Tale definizione costituisce il paradigma frequentista o statistico di assegnazione delle probabilità ad eventi; è molto usata nelle situazioni in cui sono disponibili molti dati sperimentali o osservazionali, ottenuti da replicazioni indipendenti e in condizioni costanti di un medesimo esperimento.
- Esempio Si voglia calcolare la probabilità che un neonato sia femmina. Su 100.000 nascite si sono avute 48.500 femmine. Essendo il numero di prove sufficientemente elevato, e ogni prova indipendente dall'altra, utilizziamo la definizione frequentista.
- P(F) = $\frac{48500}{100000} = 0.485$ P(M) = $\frac{51500}{100000} = 0.515$

2-Definizione frequentistica.

- Es.: Qual'è la Prob. di morte post-operatoria dopo intervento aortocoronoracico? Poiché dal 1997 al 2006 si sono registrati 24 morti su 800 interventi, la Freq. Relativa = 24/800=0,03=3% la Prob. cercata.
- La valutazione frequentista sarà tanto migliore quanto più elevato è il numero di replicazioni fisiche dell'esperimento considerato e quanto più sono rispettate le condizioni di indipendenza e di costanza delle condizioni. Naturalmente la numerosità elevata e il rispetto delle condizioni possono costituire un limite all'applicabilità di tale valutazione: si pensi ad una partita di calcio! E' necessario, allora, far riferimento ad una concezione che possa contemplare una casistica maggiore, ad esempio la definizione soggettiva.

3-Definizione soggettiva.

- **Definizione soggettiva (nelle scommesse):** la probabilità di un evento P(E) esprime il grado di fiducia che un individuo attribuisce, sulla base delle sue opinioni ed informazioni, all'avverarsi di un certo risultato (l'incertezza può riguardare un evento futuro, ma anche un evento passato) e dipende dall'opinione soggettiva, dall'ignoranza o dalla parziale conoscenza e da quante informazioni ha chi deve valutare il verificarsi (o se si è verificato) di un evento.
- E' una sintesi soggettiva di tutte le informazioni, passate e future, che si hanno sul fenomeno (es. Prob. che domani piova, che un paziente guarisca, che il Milan vinca lo scudetto, etc). Essa <u>non</u> richiede :
  - **la conoscenza del meccanismo che regola il fenomeno;**
  - **la ripetibilità del fenomeno.**

3-Definizione soggettiva.

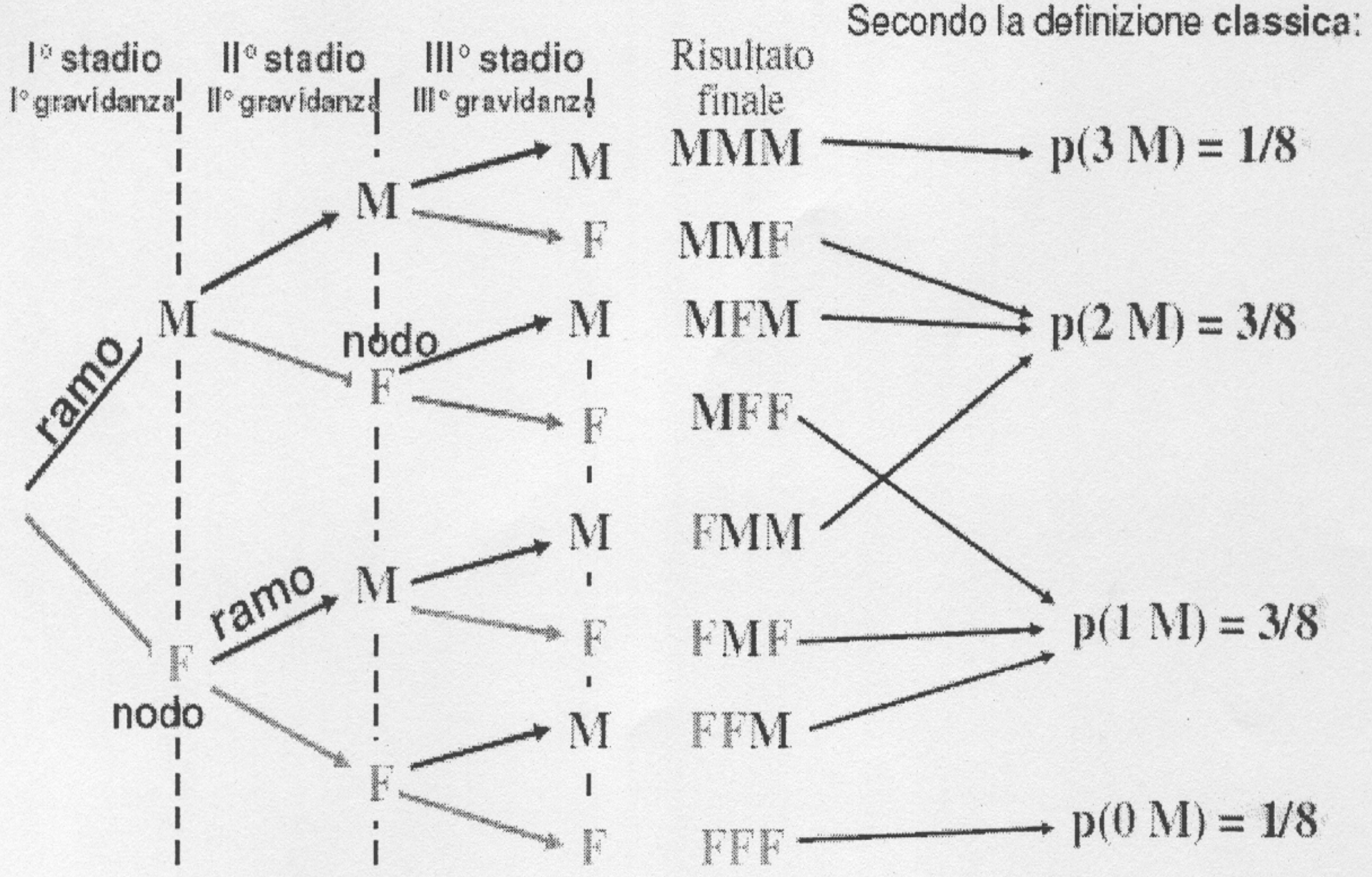
- La probabilità di un miglioramento di un paziente affetto da malattia è rappresentata dal rapporto tra l'onere che il paziente è disposto a sostenere (sacrifici e sofferenze nel sottoporsi a trattamenti terapeutici, ricoveri ospedalieri, interventi chirurgici, etc) e i benefici che ne potrà trarre in termini di miglioramento delle sue condizioni di salute. Sono disposto a pagare 100 per ottenere un "miglioramento" che, soggettivamente, valuto 200. Se un altro soggetto valuta 50 lo stesso miglioramento non sarà disposto a pagare 100.
- La combinazione delle terapie sulla quale punta il medico per ottenere un miglioramento del paziente, non può essere tale da assicurare né la guarigione certa, né uno stato permanente di malattia. Quindi, l'esito di un trattamento è sempre incerto per il medico, per cui la probabilità di questi eventi, risultati, etc sarà

$$P(E) \neq 0 \text{ e } \neq 1$$

Diagramma ad albero

Se un esperimento è a più stadi, il problema di descrivere i possibili risultati può essere semplificato mediante l'uso di diagrammi ad albero.

Esempio: *Quanti figli maschi possono nascere su 3 gravidanze?*



Assiomi del CdP

- Le definizioni date comportano concezioni diverse della probabilità di un evento, ma per tutte risultano validi i seguenti assunti (o risultati):
- 1) per ogni evento E, la probabilità P(E) è tale che:

  $$0 \le P(E) \le 1;$$

- 2) la probabilità dell’evento certo Ω è pari a:

  $$P(\Omega) = 1;$$

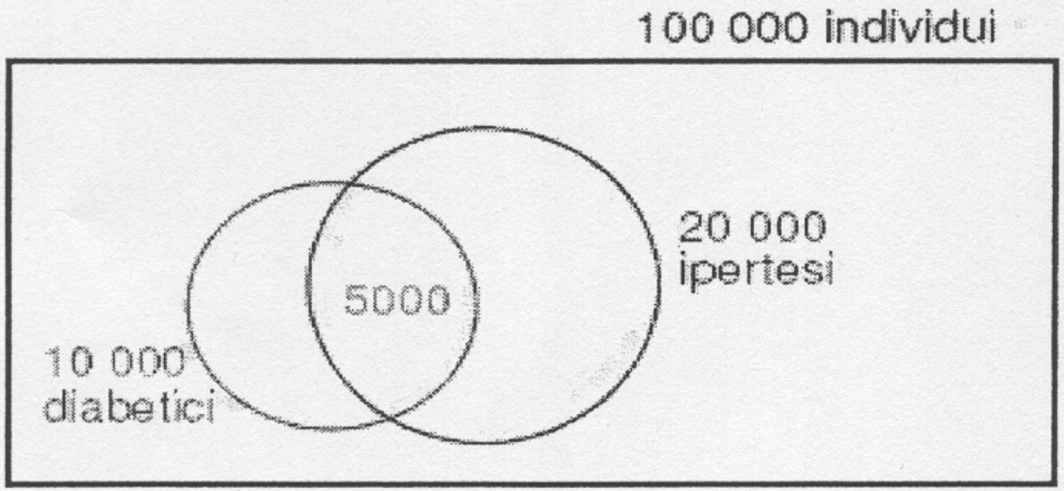
- 3) per due eventi incompatibili E1 ed E2 risulta:

  $$P(E1 \text{ o } E2) = P(E1) + P(E2).$$

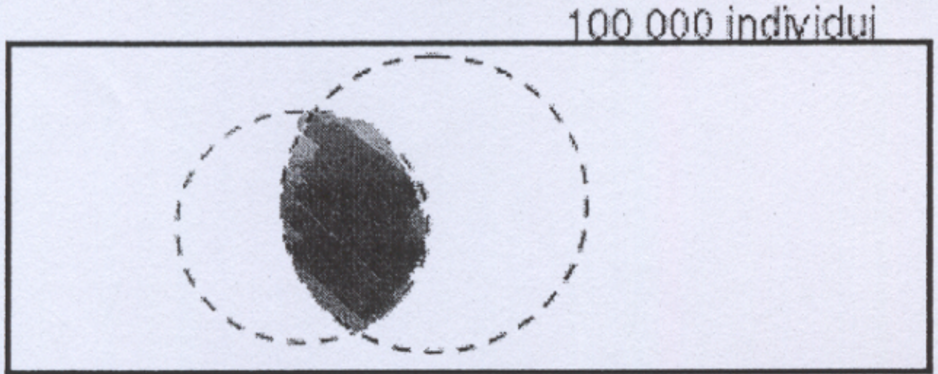
- Dalle 2) e 3) segue che la probabilità dell’evento impossibile Ø è uguale a 0 e che la probabilità dell’evento contrario $\bar{E}$ di E è dato da $P(\bar{E}) = 1-P(E)$.

ESERCIZIO: CALCOLO DELLE PROBABILITA'

In una popolazione di 100 000 individui vi sono:
10 000 diabetici (e 90 000 non-diabetici)
20 000 ipertesi (e 80 000 non-ipertesi).
5000 persone che hanno sia il diabete che l'ipertensione.

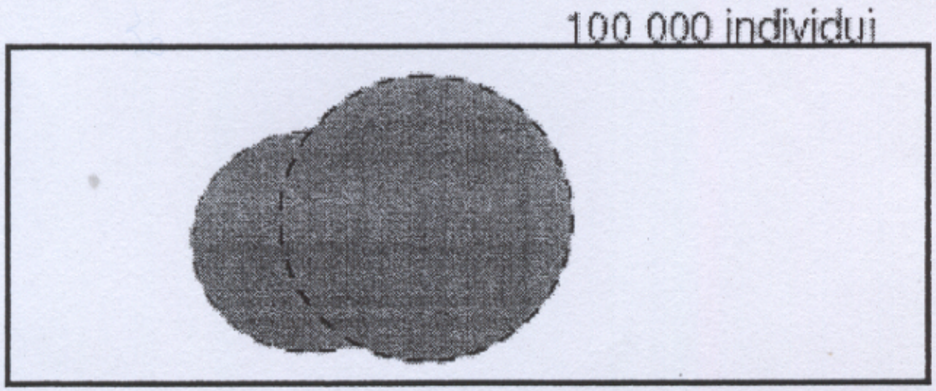


Qual è la probabilità di avere il diabete e l'ipertensione (sia il diabete che l'ipertensione)?



p (diabete ∩ ipertensione) = 5 000 / 100 000 = 0,05 = 5%

Qual è la probabilità di avere il diabete e/o l'ipertensione (solo il diabete o solo l'ipertensione o entrambi)?

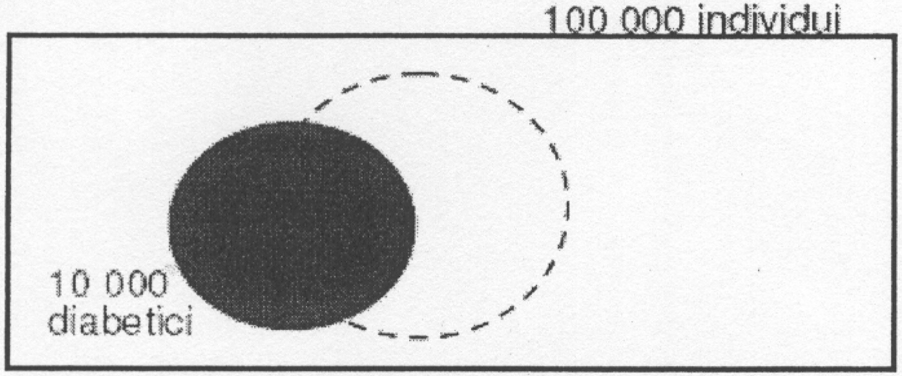


p(diabete U ipertensione) = (10000+20000-5000)/100000 = 25000/100000 =0,25=25

p(diabete U ipertensione) = p(diabete) + p(ipertensone) - p(diabete ∩ ipertensione

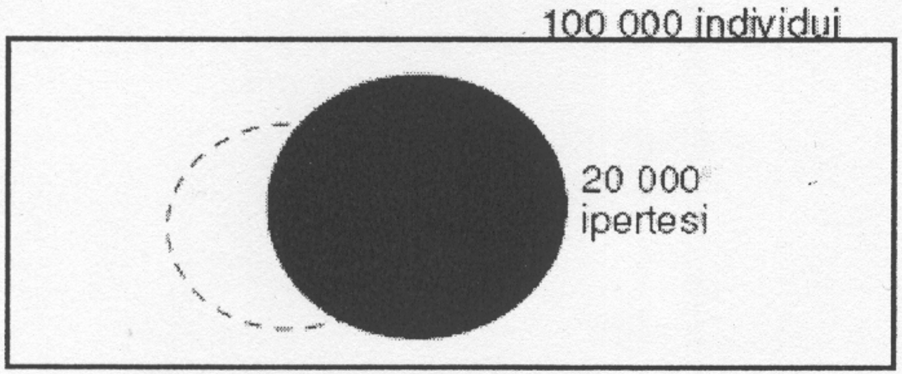
= 10% + 20% - 5% = 25%

Qual è la probabilità di avere il diabete in quella popolazione?



p (diabete) = 10 000 / 100 000 = 0,1 = 10%

Qual è la probabilità di avere l'ipertensione in quella popolazione?



p (ipertensione) = 20 000 / 100 000 = 0,2 = 20%

N.B. E' stato usato l'approccio frequentista: la probabilità è stata stimata dalla frequenza relativa

Probabilità condizionata (condizionale).

- La probabilità condizionata indica la probabilità che l'evento B si verifichi dato che si è verificato l'evento A; ossia si è interessati a quegli eventi B che si possono verificare quando si è già verificato A.
- A: evento condizionante; B: evento condizionato.
- La probabilità che io vada in montagna (M) se c'è la neve (N): prima deve essersi verificata la nevicata e poi esamino la mia andata in montagna; (nel caso che non ci sia stata nevicata, non mi pongo neanche il problema di valutare la possibilità di andare in montagna).

Esempio.

Tab.1 – Cross-classification di 100 pazienti anziani raggruppati per sesso e età

	Età			
Sesso	**A1** (70-79aa)	**A2** (80-89aa)	**A3** (≥90aa)	**Tot.**
B1 (Maschio)	**20**	**10**	**10**	**40**
B2 (Femmina)	**30**	**20**	**10**	**60**
Totale	**50**	**30**	**20**	**100**

- Selezionando un individuo casualmente, qual'è la probabilità che sia un Maschio di età 70-79 anni?

$P(A1 \cap B1) = 20/100$

Data l'indipendenza tra le due variabili età e sesso

- Qual è la probabilità che sia una Femmina di età ≥ 90 anni?

$Pr\ (A3 \cap B2) = 10/100$

- Queste sono **probabilità congiunte**.

Probabilità marginale.

- Data una cross-classification di individui secondo una o più caratteristiche (TAB.1), la probabilità associata con una di queste caratteristiche, senza considerare le altre, è chiamata probabilità marginale.
- Scegliendo un individuo in modo random. Qual'è la probabilità che sia un Maschio?
- **Pr (B1)** = Pr (B1∩ A1) + Pr (B1 ∩ A2) + Pr (B1∩ A3) = 20/100 + 10/100
- +10/100 = 40/100 = 0.4 = 40%
- Qual è la probabilità che sia un individuo di età compresa tra 80 e 89 anni?
- **Pr (A2)** = Pr (A2 ∩ B1) + Pr (A2 ∩ B2) = 10/100 + 20/100 = 30/100

Probabilità condizionata.

- In generale se A e B sono <u>indipendenti</u>

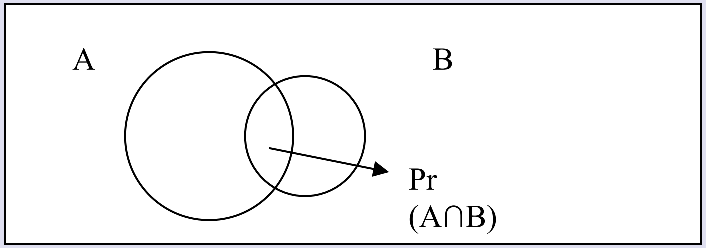
Pr (B/A) = $\frac{\Pr(B \cap A)}{\Pr(A)} = \frac{\Pr(B)\Pr(A)}{\Pr(A)} = \Pr(B)$

- Pr (B1/A3) “Probabilità di estrarre un maschio, dato che è stato estratto un soggetto con età ≥90 anni”

Pr (B1/A3) = $\frac{\Pr(B1 \cap A3)}{\Pr(A3)} = \frac{10/100}{20/100} = \frac{10}{20}$

- Poichè la Pr (B1/A3) ≠ Pr (B1) estrarre un maschio >90 anni è diversa da estrarre un M sul totale

$$\frac{10}{20} \neq \frac{40}{100}$$



Esempio di eventi indipendenti.

- Data la seguente tabella:

	A	$\overline{\text{A}}$	Totale
	Maschio	Femmina	
B ipertensione	10 (20%)	20 (20%)	30
$\overline{\text{B}}$ normotensione	40	80	120
Totale	50	100	150

- Pr (B/A): qual è la probabilità di estrarre un paziente iperteso dato che è Maschio?
- Pr (B/A) = $\frac{\Pr(A \cap B)}{\Pr(A)} = \frac{10/150}{50/150} = \frac{10}{50}$ =20%
- Prob = 20% ……….a che sia Femmina

Esempio di eventi indipendenti.

- Pr (B) = Pr (ipertensione) = $\frac{30}{150} = \frac{10}{50}$
- Quindi Pr (B/A) = Pr (B)
- Qual è la probabilità di estrarre un paziente iperteso dato che è Femmina?
- Pr (B/$\bar{A}$) = $\frac{\Pr(\bar{A} \cap B)}{\Pr(\bar{A})} = \frac{20/150}{100/150} = \frac{20}{100} = \frac{10}{50} = \Pr(\mathrm{B})$
- Gli eventi A e B sono indipendenti (l'ipertensione non dipende dal sesso).
- All'informazione che una persona sia Maschio o Femmina non consegue alcuna informazione riguardo la sua pressione sanguigna.
- Nota: Pr (A ∩ B) = 10/150 = Pr (A) Pr (B) = $\frac{50}{150} \cdot \frac{30}{150} = \frac{10}{150}$

Esempio di eventi indipendenti.

- In Statistica il concetto di “indipendenza” è cruciale quando valutiamo i campioni, piuttosto che le popolazioni.

Malattia	Esposizione	
	Si (E)	No ($\overline{E}$)
Presente (D)	A	B
Assente ($\overline{D}$)	C	D

- Se la malattia è “indipendente” dall’esposizione ad un fattore di rischio

  $Pr\ (D/E) = Pr\ (D)$; $Pr\ (D/\overline{E}) = Pr\ (D)$; $Pr\ (D/E) = Pr\ (D/\overline{E})$

- E la malattia non è in relazione, o non risulta associata, con l’esposizione.
- Le Probabilità Condizionate assumono particolare interesse in Medicina: per i valori predittivi positivi o negativi di un test diagnostico, per la diagnosi, per l’analisi della sopravvivenza, delle ricadute, etc

Teorema della probabilità composta.

- Per due eventi E1 e E2 **indipendenti** la probabilità dell'evento:

$$P(E1 \text{ e } E2) = P(E1) \cdot P(E2);$$

- Per eventi E1 e E2 **dipendenti** la probabilità dell'evento:

$$P(E1 \text{ e } E2) = P(E1) \cdot P(E2/ E1)$$

Teorema della probabilità totale.

- Se due eventi E1 e E2 sono **incompatibili** la probabilità dell'evento:

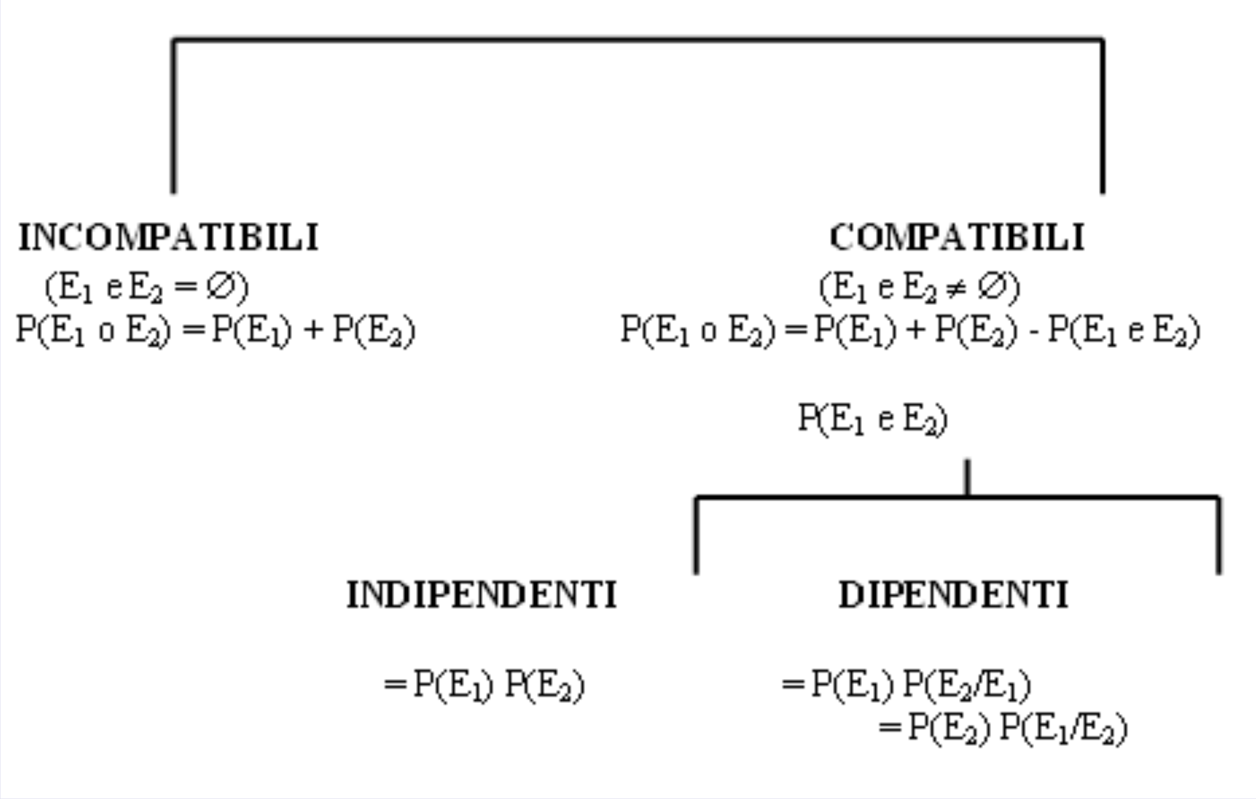
$$P(E1 \text{ o } E2) = P(E1) + P(E2),$$

- Se invece gli eventi E1 e E2 sono **compatibili,** allora:

$$P(E1 \text{ o } E2) = P(E1) + P(E2) - P(E1 \text{ e } E2)$$

Riassumendo:

- Gli EVENTI **E1 e E2 possono** essere:



Conclusione:

- Mediante i due teoremi del CdP (Probabilità Composte e Totali) si possono calcolare le probabilità di eventi complessi dopo averli:
    1. disaggregati in eventi semplici;
    2. individuato il tipo di connettivo logico che li lega;
    3. verificato la incompatibilità o la compatibilità e l’indipendenza o la dipendenza degli eventi semplici.

Esempio (vedi diapositiva 10):

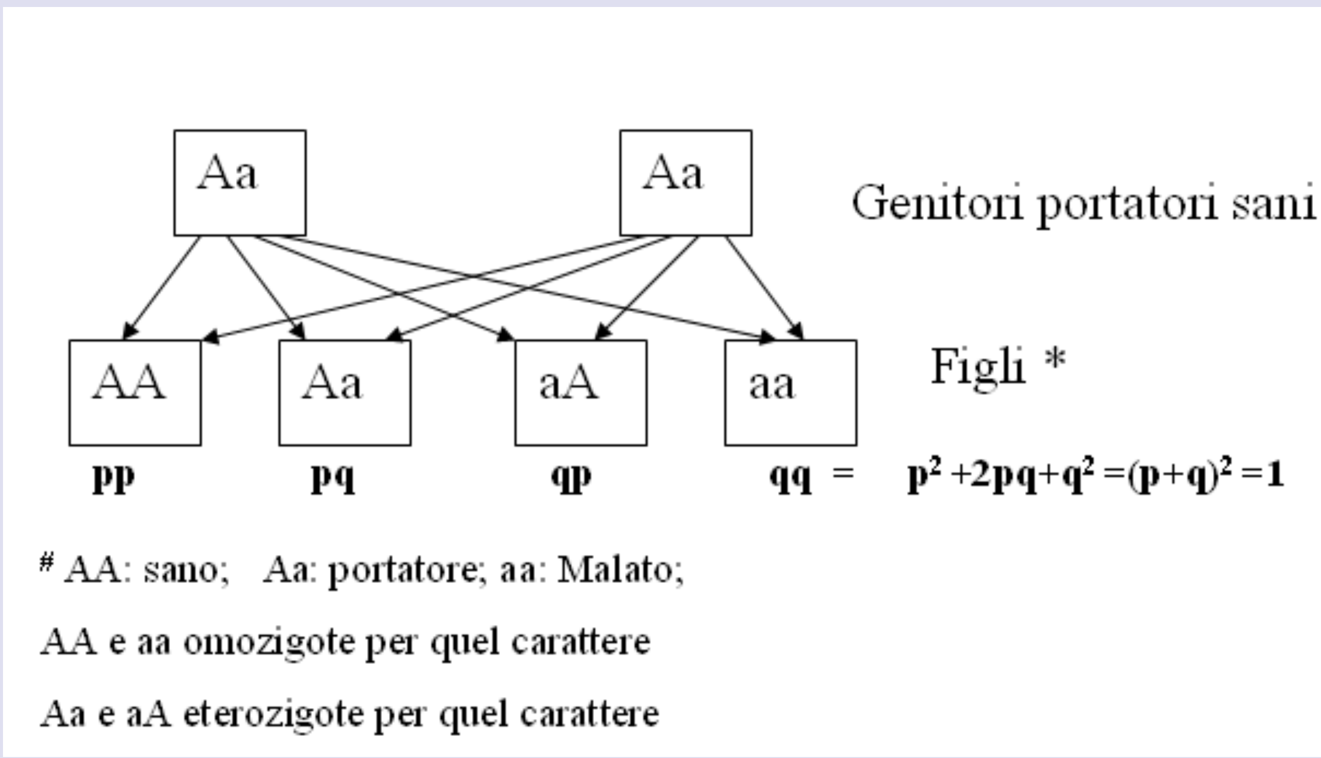
- Sapendo che per un generico individuo la probabilità di essere miope è di 0.20, quella di essere mancino è di 0.08, e che le due condizioni sono **indipendenti,** si chiede di valutare la probabilità e il numero di individui che un gruppo di 1000 persone siano:
  a) miopi e mancini;
  b) miopi o mancini.
- Consideriamo gli eventi: E1 = “miope”; E2 = “mancino”
- **P(E1 e E2)** = P(E1)·P(E2) = 0,20·0,08 = 0,016
- **P(E1 o E2)** = P(E1)+P(E2)-P(E1 e E2) = 0,20 + 0,08- 0,016 = 0,264.
- Ipotizzando allora una popolazione di 1000 studenti, ci si aspetta di trovare:
- 0.016·1000 = 16 soggetti sia miopi che mancini,
- 0.264·1000 = 264 individui o miopi o mancini.

Applicazioni alla genetica.

- **L**EGGE DI INDIPENDENZA: nella trasmissione di 2 diversi caratteri, le possibili combinazioni relative al primo carattere sono indipendenti da quelle relative al secondo carattere.
- **Esempio** Malattia dovuta ad un Allele recessivo.
- Si tratta di studiare da un punto di vista statistico il caso dei geni con due soli alleli, che usualmente vengono indicati con una stessa lettera, maiuscola e minuscola, di solito **A** e **a**.
- **Il genotipo** di un individuo è caratterizzato da 4 possibili coppie di alleli * che in realtà sono 3 # (Aa;aA; indistinguibili)

Applicazioni alla genetica.

- Se i genitori sono entrambi portatori sani del gene della talassemia, la Prob di avere un figlio malato è =1/4 = 25% e 2/4=50% la prob che sia portatore sano e il restante 25% risulterà sano.



Applicazioni alla genetica.

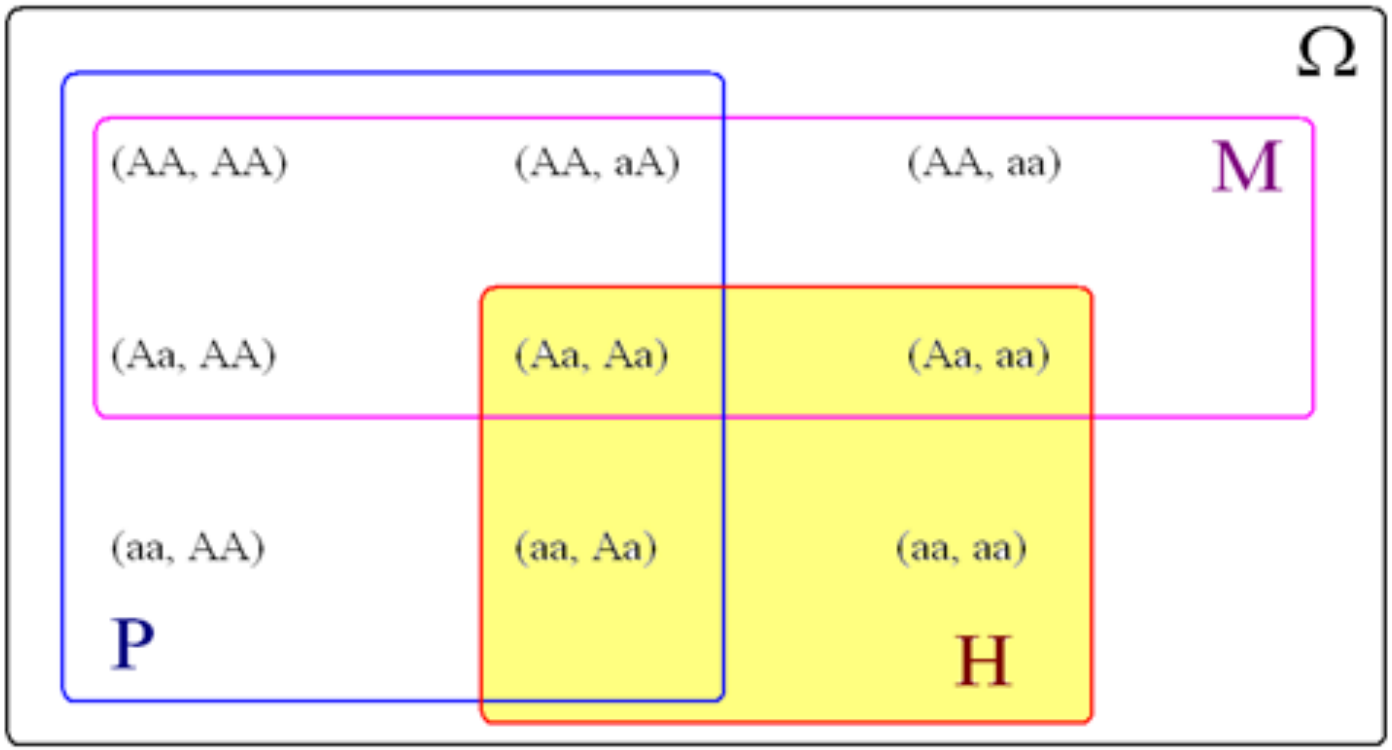
- Quanto al **fenotipo,** si usa parlare di individui di tipo A e di tipo a:
- il gene A è dominante rispetto ad a, se gli eterozigoti Aa hanno lo stesso aspetto degli omozigoti AA, e il gene a è recessivo rispetto ad A. In questa situazione gli individui A possono essere AA oppure Aa, mentre quelli a sono tutti omozigoti aa.
- La trasmissione ereditaria di ciascun carattere avviene con un <u>meccanismo di scelta casuale</u> di uno dei due alleli del padre (P) e di uno dei 2 della madre (M), che si ricompongono a formare una nuova coppia di alleli nei cromosomi del figlio (F).

Applicazioni alla genetica.

- Così se la probabilità di scelta di **A** è **p e** quella di **a** è **q=1- p**, risulta:
- P(AA) = **pp**, P(Aa) = P(aA)=**pq** e P(aa) = **qq.**
- E si ha: $\mathbf{pp + pq + qp + qq = p^2 + 2pq + q^2 = (p+q)^2 = 1}$
- Considerata una malattia dovuta alla presenza di un allele recessivo, gli eventi si possono rappresentare come in figura seguente dove:

  P = il padre è sano; M = la madre è sana,

Applicazioni alla genetica.



Applicazioni alla genetica

- Mentre l’evento **H** individua tutte quelle coppie padre-madre che possono eventualmente generare un figlio malato
- **Legge di disgiunzione:** se un genitore è Aa e l’altro omozigote (AA o aa), le due combinazioni AA ed Aa, ossia **Aa ed aa,** sono equiprobabili. Se i genitori sono eterozigoti, sono equiprobabili le 4 combinazioni **AA, Aa, aA, aa**.
- P(AA)=P(aa)=1/4 e P(Aa)=1/2.

Esempio: matrimonio tra consanguinei.

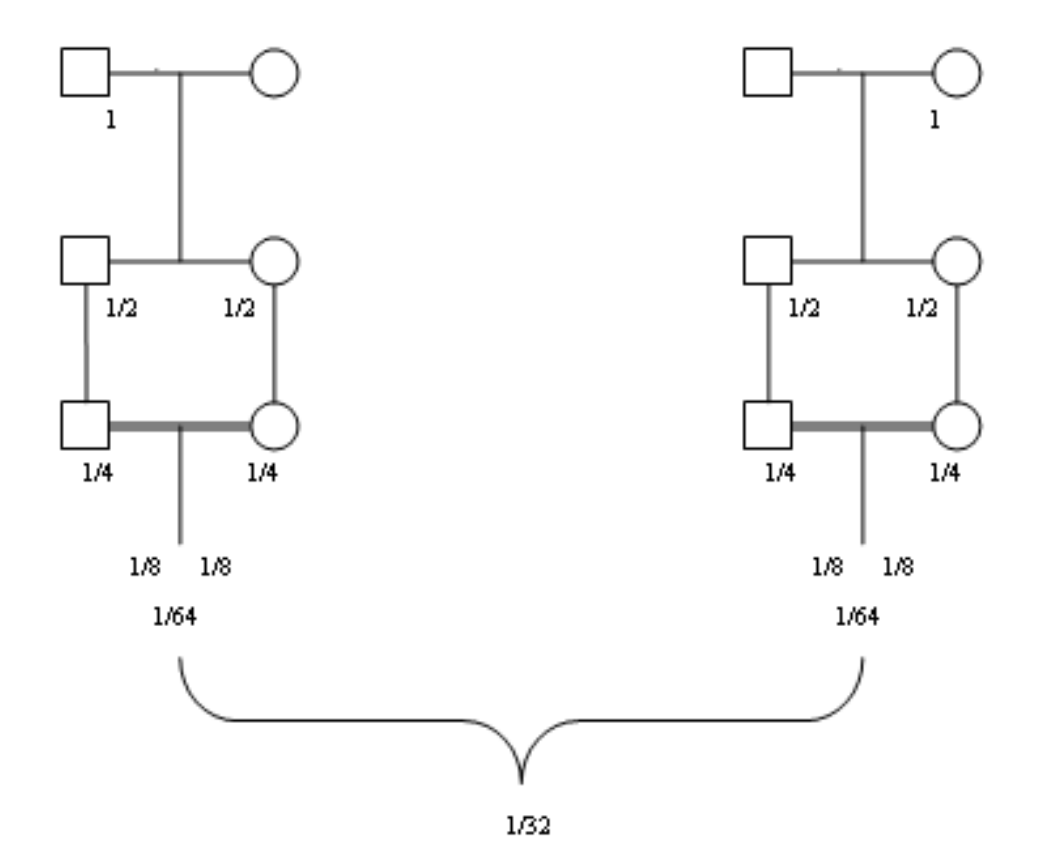
- L'unione tra cugini primi rappresenta una delle più frequenti richieste nella consulenza genetica. Questo tipo di matrimonio è legale in molti paesi occidentali, anche se può essere oggetto di restrizioni religiose o sociali; in alcune comunità asiatiche o nordafricane i matrimoni tra cugini primi sono favoriti.
- La stima del rischio genetico per i figli nati da matrimoni tra consanguinei, quando nella famiglia non esistono soggetti affetti da particolari malattie, dipende dal numero di geni recessivi mutati presenti negli individui sani della popolazione generale.

Esempio: matrimonio tra consanguinei.

- Numerosi studi hanno dimostrato che un individuo è portatore di almeno un gene mutato per una data malattia recessiva e probabilmente di almeno due geni per malattie letali responsabili di aborti spontanei o di nati morti.
- Nel caso di un matrimonio tra cugini primi il rischio che un gene recessivo dannoso arrivi alla prole per discendenza di entrambi i genitori è di 1/64. Dato che, però, bisogna considerare anche l’altro antenato comune, il rischio risulta raddoppiato (rischio totale = 1/32).

Esempio: matrimonio tra consanguinei.

- Se si applica lo stesso tipo di ragionamento al caso di due geni letali presenti in entrambi i coniugi, il rischio sale a 1/16.



Esempio di genetica di popolazioni:

- La legge di Hardy-Weinberg.
- Si tratta di un modello probabilistico che consente di calcolare le frequenze genotipiche ad un locus. Tale regola è valida per popolazioni numerose e quando gli effetti della mutazione, della selezione e della migrazione sono piccoli. E’ necessario, inoltre, che vi sia accoppiamento casuale per il carattere di riferimento. Per moltissimi loci tali condizioni sono verificate e quindi la Legge di Hardy-Weinberg, per il calcolo delle frequenze genotipiche, si rivela di grande utilità.

Esempio di genetica di popolazioni:

- Indicata con: p=frequenza dell’allele dominante e con q=frequenza dell’allele recessivo.

- L’equilibrio è espresso dalla seguente uguaglianza probabilistica:

  $(p+q)2=p2+2pq+q2=1$ (distribuzione dei genotipi all’equilibrio)

Esempio di genetica di popolazioni:

- Si può utilizzare come esempio quello del gruppo sanguigno MN (in questo caso i due alleli M e N sono codominanti). In uno studio condotto su 425 soggetti (genotipi osservati), si sono trovati:

  137 genotipi MM 207 genotipi MN 81 genotipi NN

- Si vogliono trovare i rapporti tra le frequenze genotipiche e le frequenze dei singoli alleli M e N. La frequenza all'elica degli individui considerati è di 850 nel locus MN (425x2=850). Gli omozigoti MM contribuiscono con 137x2=274 alleli, gli eterozigoti MN con 207 alleli M e 207 alleli N, e gli omozigoti NN con 81x2=162 alleli N.

Esempio di genetica di popolazioni:

- Tabella 1. Esempio di calcolo delle frequenze geniche per il locus MN.

Genotipo	MM	MN	NN
Numero di soggetti	137	207	81
Numero di alleli del locus	274 M	207 M + 207 N	162 N
Totale alleli M (o N)	481 M 369 N		

- Complessivamente ci sono 481 alleli M e 369 alleli N su un totale di 850, quindi:
- frequenza allelica di M = 481/850 = 0.5659 = p
- frequenza allelica di N = 369/850 = 0.4341 = q

Il teorema di Bayes.

- Il Teorema di Bayes permette di ricavare la probabilità di un dato evento (H), avendo osservato un fatto (evento E) eventualmente “originato” da H.
- L’evento E non è necessariamente la conseguenza diretta ed esaustiva di H, potrebbero esistere, infatti, altre "cause" che avrebbero potuto generare lo stesso evento.

Il teorema di Bayes.

- Esempio:
- Grossa vincita ad un gioco (H) → il vincitore diventa miliardario (E)
- Ma H non è una causa esclusiva di E. Se si osserva un soggetto miliardario (E) si può solo dire che è PROBABILE che H (la vincita) ne sia stata la causa.
- Inoltre data una causa H possono seguire diversi effetti (E_i).
- In generale, di fronte a diverse cause (H_i) e diversi effetti (E_i) non si può risalire, con certezza, dagli E_i alle H_i generatrici.

Il teorema di Bayes.

- <u>Esempio 1:</u>
- Da una coppia nascono 5 figlie femmine (E) la “causa” potrebbe essere attribuita alla tendenza della coppia a generare figlie femmine (H).
- Ipotizzando $P(F) \cong P(M) \cong 1/2$ l'evento verificatosi ha probabilità:
- $P(E)=(1/2)^5=1/32=0.03$ che risulta molto bassa.
- Il numero di figlie porterebbe a ritenere che quei genitori tendano a generare femmine, ma la figliolanza di 5 femmine è un evento raro (P(E)=0.03) dovuto al “CASO”. L’erroneità della precedente deduzione sta nel confondere la PROBABILITA’ che si verifichi un E con la PROBABILITA’ INVERSA (ignota) che, essendosi verificato un fatto (E), esso dipende dalla causa H (tendenza a generare figlie femmine), invece si è prodotto per effetto del caso, anche se la sua probabilità è molto piccola (3%).

Il teorema di Bayes.

- Tutta questa problematica è stata formalizzata da Bayes circa 2 secoli fa (1763).
- Il **teorema di Bayes** permette di determinare, data l'osservazione di un dato evento E, la probabilità degli eventi-cause H_1, H_2, …, H_i,…, H_n necessari ed incompatibili.
- Le **probabilità** $P(H_1)$, P(H2), …, $P(H_i)$, …, $P(H_n)$, dette **a priori,** sono stimabili dalla letteratura (ad esempio epidemiologica).
- Le **probabilità** $P(E\backslash H_1)$,…, $P(E\backslash H_i)$, …, $P(E\backslash H_n)$, dette **probative**, sono date dalla letteratura scientifica e stimano la probabilità di E condizionata al verificarsi di H_i
- La PROBABILITA' A POSTERIORI risponde alla formula (Bayes)

$$P(Hi/E) = \frac{P\,(Hi)\,P\,(E/Hi)}{P(Hi)P(E(/Hi) + ... + ...P(Hi)P(E/Hi) + ... + P(Hn)P(E/Hn)}$$

- probabilità che sia H_i la "causa" della realizzazione di E. Calcolando $P(H_i/E)$, per ciascun H_i, viene presa quale "causa" di E l'evento H_i per il quale $P(H_i/E)$ è più alta.

Il teorema di Bayes.

- Esempio 2:
- I gemelli possono essere veri gemelli (VG) (omozigoti) e sono dello stesso sesso, pseudo-gemelli (PG) (eterozigoti) e, in tal caso, è 1/2 la probabilità che siano dello stesso sesso.
- Consideriamo gli eventi:

E = stesso sesso ("i gemelli sono dello stesso sesso",)

E1 = omozigoti/e (VG = "i gemelli sono veri gemelli"),

E2 = eterozigoti/e (PG = "i gemelli sono pseudo-gemelli").

- Probabilità a priori P(VG)=0.35 P(PG)=0.65
- Probabilità probative P(E/VG)=1 P(E/PG)=0.5

Il teorema di Bayes.

- Probabilità a posteriori: che dati due gemelli dello stesso sesso, siano omozigoti

$$P(VG/E) = \frac{P(VG)\,P(E/VG)}{P(VG)\,P(E/VG) + P(PG)\,P(E/PG)} = \frac{0.35 \cdot 1}{0.35 \cdot 1 + 0.65 \cdot 0.5} = 0.52;$$

$$P(PG/E) = \frac{P(PG)\,P(E/PG)}{P(VG)\,P(E/VG) + P(PG)\,P(E/PG)} = \frac{0.65 \cdot 0.5}{0.35 \cdot 1 + 0.65 \cdot 0.5} = 0.48$$

- 0.52 è la probabilità che avendo riscontrato due gemelli dello stesso sesso questi siano omozigoti, e 0.48 che siano eterozigoti.

Il teorema di Bayes.

- <u>Esempio 3:</u>
- E: l'infarto del miocardio
- C1 fumo da sigaretta, C2 ipertensione, C3 il caso (le altre cause)
- Prob a priori P(C1)=0.20, P(C2)=0.40, P(C3)=0.40
- Prob probative P(E/C1)=0.30 , P(E/C2)=0.80 , P(E/C3)=0.50
- Se E si è verificato con quale probabilità ha agito Ci?
- PROBABILITA' A POSTERIORI (Bayes)
- La probabilità di essere fumatore, se ha avuto l'infarto, è data da:

$$P(C_1 / E) = \frac{0.20 \cdot 0.30}{0.20 \cdot 0.30 + 0.40 \cdot 0.80 + 0.40 \cdot 0.50} = 0.10$$

- La probabilità di essere iperteso, se ha avuto l'infarto, è data da:

$$P(C_2 / E) = \frac{0.40 \cdot 0.80}{0.20 \cdot 0.30 + 0.40 \cdot 0.80 + 0.40 \cdot 0.50} = 0.55$$

Il teorema di Bayes.

- Preso a caso un paziente infartuato, la probabilità che egli non sia né iperteso né fumatore (che abbiano agito altre cause, il caso) è data da:

$$P(C_3 / E) = \frac{0.40 \cdot 0.50}{0.20 \cdot 0.30 + 0.40 \cdot 0.80 + 0.40 \cdot 0.50} = 0.34$$

- allora la causa più probabile è C2 (l'ipertensione).

Il teorema di Bayes.

- **Esempio 4**
- Supponiamo che un paziente abbia l'ulcera duodenale (E).
- Con quale probabilità hanno agito le seguenti cause?
- C1: presenza di Helicobacter pilory (Hp^+)
- C2: Abitudini di vita non razionali (tipo di alimentazione, fumo, etc)
- C3: Il caso (le altre cause)
- Prob. a priori: P(C1)=0.50, P(C2)=0.30, P(C3)=0.20.
- Prob. probative: P(E/C1)=0.70, P(E/C2)=0.20, P(E/C3)=0.10
- La probabilità che un paziente avendo l'ulcera essa dipenda dalla presenza dell'Hp^+ è data da:

$$P(C_1 / E) = \frac{0.50 \cdot 0.70}{0.50 \cdot 0.70 + 0.30 \cdot 0.20 + 0.20 \cdot 0.10} = 0.81$$

Il teorema di Bayes.

- La probabilità che un paziente avendo l'ulcera essa dipenda dall'igiene di vita scadente è data da:

$$P(C_2 / E) = \frac{0.30 \cdot 0.20}{0.50 \cdot 0.70 + 0.30 \cdot 0.20 + 0.20 \cdot 0.10} = 0.14$$

- la probabilità che un paziente avendo l'ulcera essa dipenda da altre cause è data da:

$$P(C_3 / E) = \frac{0.20 \cdot 0.10}{0.50 \cdot 0.70 + 0.30 \cdot 0.20 + 0.20 \cdot 0.10} = 0.05$$

- E' molto probabile (81%), quindi, che l'ulcera sia stata causata dall'Hp^+.