

ESERCIZI DI CALCOLO DELLE PROBABILITA'

ES 1

Supponiamo che una certa forma di allergia respiratoria colpisca di norma 1 individuo ogni 20, mentre le intolleranze alimentari riguardano il 3.5% dei casi. Supponendo che i due eventi siano indipendenti, qual è la probabilità di avere entrambi i problemi? e di averne almeno uno? e di averne solo uno?

ES 2

Da uno studio sul Body Mass Index effettuato in una popolazione, si stima che il 33% degli individui è Normopeso, il 50% Sovrappeso, e il 17% Obeso. In questi 3 gruppi, la probabilità di sviluppare una certa tipologia di malattie cardiovascolari è rispettivamente pari a 1%, 3% e 6%. Sapendo che la popolazione comprende complessivamente 10,000 individui, calcolare:

- Quanti soggetti obesi dovrebbero complessivamente esserci in questa popolazione;
- Qual è la probabilità, estraendo a caso un individuo dalla popolazione, che sia un soggetto Normopeso e si ammali di una di queste malattie;
- Qual è la probabilità che un individuo scelto a caso in questa popolazione sviluppi una di queste malattie – e, quindi, quanti casi di malattia ci aspettiamo nella popolazione.

ES 3

Nell'anemia mediterranea, un genitore portatore sano (microcitemico) trasmette l'anomalia genetica al figlio con probabilità del 50%; se l'individuo eredita l'anomalia da entrambi i genitori, egli si ammala di talassemia; se la eredita solo da un genitore, egli è a sua volta un portatore sano.

- Qual è la probabilità che il figlio di due portatori sani sia malato? e la probabilità che sia un portatore sano? e che non erediti l'anomalia?
- A sua volta, il soggetto in questione (figlio di due portatori sani; chiamiamolo Tizio) potrà avere in futuro un figlio; immaginando che Tizio si accoppi con un soggetto che non presenta l'anomalia, con che probabilità il figlio di Tizio (ossia, il nipote della prima coppia di genitori) sarà portatore sano?

Suggerimento per l'impostazione: abbiamo 3 possibili situazioni, con Tizio portatore sano / malato di talassemia / senza anomalia, e la probabilità che il figlio di Tizio sia portatore sano dipende dalla situazione ...

ES 4

Nel concepimento, per una certa tipologia di coppia, la probabilità che il feto presenti una certa malformazione congenita è pari a 15%; effettuando in gravidanza un test che individua con certezza la presenza di questa malformazione su 8 donne, con che probabilità si troveranno 2 casi di malformazione? e nessuno?

ES 5

Se la mortalità per influenza A H1N1 è pari a 1 caso di morte ogni 10,000 malati, qual è la probabilità che fra 100 malati non ne muoia nessuno?

ES 6

Se il peso alla nascita dei neonati maschi è distribuito secondo una Normale di media 2.8 e varianza 0.16, qual è la probabilità che nasca un bimbo che pesi meno di 1.8 kg?.

ES 7

Se la prevalenza della microcitemia in una regione è del 15%, somministrando a 100 bambini un test che individua la presenza di microcitemia con sensibilità 80% e specificità 98, quanti test positivi dovremmo avere?

ES 8

Lo Skin-prick-test è utilizzato frequentemente per lo screening delle allergie agli alimenti; si tratta di un test cutaneo che fornisce il risultato in 15 minuti, ha sensibilità pari ad almeno 90%, ma ha una scarsa specificità, in media pari a 60%.

Qual'è il problema principale di questo test? Dire se ciascuna delle seguenti affermazioni è Vera o Falsa.

- a) Indica troppi falsi negativi
- b) Indica troppi falsi positivi
- c) Conduce spesso ad una diagnosi errata di allergia alimentare, mentre il soggetto non è allergico
- d) E' molto probabile che un soggetto allergico risulti negativo al test

Secondo alcuni studi, le allergie alimentari si osservano nel 6% circa dei bambini e nell'1.5% degli adulti. Dire se ciascuna delle seguenti affermazioni è Vera o Falsa e, se Falsa, suggerire una correzione alle parti sottolineate.

- e) Un bambino sottoposto allo Skin-prick-test ha test positivo con probabilità 6%.
- f) Un bambino sottoposto allo Skin-prick-test ha test positivo con probabilità 90%.
- g) Un adulto allergico sottoposto allo Skin-prick-test ha test positivo con probabilità 90%.
- h) Un bambino con Skin-prick-test positivo ha probabilità di essere allergico pari a 13%.
- i) La probabilità di essere allergico quando lo Skin-prick-test è positivo è la stessa per adulti e bambini

SOLUZIONI

ES 1 (allergia e intolleranze)

Gli eventi che stiamo considerando e i dati del problema sono:

A = Allergia Respiratoria, $p(A)=1/20=0.05$

I = Intolleranza Alimentare, $p(I)=0.035$.

A ed I eventi indipendenti

Formalizziamo le domande e risolviamo:

- Probabilità di avere entrambi: $p(A \& I)$ [intersezione] = [per la proprietà di indipendenza] = $p(A) \cdot p(I)=0.05 \cdot 0.035=0.00175$ (approssimando: circa 2 per mille)

- Probabilità di averne almeno uno: $p(A \text{ oppure } I)$ [unione] = $p(A)+p(I)-p(A \& I) = 0.05+0.035-0.00175=0.08325$ (circa 8.3%)

- Probabilità di averne uno solo: vediamo 2 procedimenti.

Il primo mostra una scomposizione dell'evento in questione, utile in generale: osserviamo che la probabilità che cerchiamo è

$p((A \& \text{non } I) \text{ oppure } (I \& \text{non } A))$ – giusto?

Osserviamo che l'unione ["oppure"] riguarda due eventi incompatibili ossia disgiunti; quindi possiamo scrivere:

$p((A \& \text{non } I) \text{ oppure } (I \& \text{non } A)) = p(A \& \text{non } I) + p(I \& \text{non } A) = (*)$

Gli eventi "non I" e "non A" sono rispettivamente i complementari di I e A, e quindi:

$p(\text{non } I)=1-0.035=0.965$; $p(\text{non } A)=1-0.05=0.95$

Essendo A ed I eventi indipendenti, sono anche indipendenti ciascuno col complementare dell'altro; quindi le probabilità delle intersezioni sono il prodotto delle probabilità; quindi:

$(*) = p(A) \cdot p(\text{non } I) + p(I) \cdot p(\text{non } A) = 0.05 \cdot 0.965 + 0.035 \cdot 0.95 = 0.04825 + 0.03325 = 0.0815$

Secondo procedimento, più intuitivo: "solo uno" vuol dire "o l'uno o l'altro ma non entrambi", il che ci suggerisce che la probabilità che cerchiamo è:

$pr(\text{l'uno o l'altro}) - pr(\text{entrambi}) = pr(\text{unione}) - pr(\text{intersezione}) = 0.08325 - 0.00175 = 0.0815$

ES 2 (Malattie cardiovascolari e BMI)

Gli eventi che stiamo considerando sono:

N=Normopeso, $p(N)=0.33$; S=Sovrappeso, $p(S)=0.5$; O=Obeso, $p(O)=0.17$. Osserviamo che questi 3 eventi sono incompatibili (disgiunti), e insieme ricostruiscono lo spazio Ω ($N \cup S \cup O = \Omega$); ovvero, osserviamo che $p(N)+p(S)+p(O)=1$.

Inoltre, abbiamo l'evento M, sviluppare una certa tipologia di malattie cardiovascolari. Il testo ci fornisce le probabilità di M nelle 3 situazioni di peso precedenti – probabilità condizionate:

$p(M|N)=0.01$ $p(M|S)=0.03$ $p(M|O)=0.06$

Ultimo dato disponibile: la popolazione comprende $n=10,000$ individui.

Passiamo alle domande:

a) numero di soggetti obesi: $p(O) \cdot n = 0.17 \cdot 10,000 = 1,700$

b) $p(N \cap M) = p(N) \cdot p(M|N) = 0.33 \cdot 0.01 = 0.0033 = 3.3 \text{ ‰}$ (permille)

c) $p(M)$: per impostare il calcolo, dobbiamo considerare che ci si ammala con probabilità diversa a seconda della condizione di peso, e conosciamo queste probabilità, e inoltre che nella popolazione ciascuna condizione ha una frequenza (prevalenza) diversa, che pure conosciamo; in pratica, partiamo dalla considerazione che: ci si ammala e si è Normopeso, oppure Sovrappeso, oppure Obesi:

$$p(M) = p((M \cap N) \cup (M \cap S) \cup (M \cap O))$$

sono eventi disgiunti (poiché N, S e O disgiunti) $\rightarrow$

$$p(M) = p(M \cap N) + p(M \cap S) + p(M \cap O)$$

la prima l'abbiamo calcolata, le altre due si calcolano in maniera analoga, quindi:

$$p(M) = p(M|N) \cdot p(N) + p(M|S) \cdot p(S) + p(M|O) \cdot p(O) = 0.0033 + 0.015 + 0.0102 = 0.0285 \approx 3\%$$

$$\rightarrow \text{numero malati} = p(M) \cdot n = 0.0285 \cdot 10,000 = 285$$

- Osserviamo che abbiamo applicato una sorta di media ponderata delle probabilità condizionate di ammalarsi, con pesi dati dalle prevalenze di ciascuna condizione (la somma dei pesi è pari a 1, come osservato all'inizio).

- Osserviamo anche che l'addendo più grande nell'ultima somma, e quindi il contributo più grande al numero complessivo di malati, viene dal gruppo dei Sovrappeso: sebbene il rischio sia molto maggiore per gli Obesi ($RR_{OvsN}=6$, $RR_{OvsS}=3$), esso si applica a una porzione più piccola della popolazione; un piccolo rischio, applicato a tanti individui, implica un grosso aggravio per la popolazione – questo genere di considerazioni è utile in ambito epidemiologico e di salute pubblica.

ES 3 (Trasmissione dell'anemia mediterranea)

Chiamiamo P l'evento "portatore sano" (microcitemico), M l'evento "malato di talassemia", e S (impropriamente, per "sano") la situazione di assenza di anomalia genetica (che corrisponde all'evento complementare all'unione di questi due).

L'altro evento di interesse è T=trasmissione dell'anomalia; il testo dice che $p(T|P)=0.50$; inoltre, come è facile immaginare, $p(T|M)=1$, $p(T|S)=0$.

a)

- l'individuo è malato se riceve l'anomalia da entrambi i genitori; la trasmissione da un genitore è indipendente da quella dall'altro genitore, quindi questa probabilità è: $p(T|P) \cdot p(T|P) = 0.25$;

- l'individuo è portatore sano se riceve l'anomalia da uno dei due genitori, ma non dall'altro; quindi questa probabilità è: $2 \cdot p(T|P) \cdot (1-p(T|P)) = 0.5$ (la moltiplicazione per 2 è riferita al fatto che può riceverla dal padre – e non dalla madre – o viceversa, dalla madre – e non dal padre);

- l'individuo è sano se non riceve l'anomalia ne' dal padre, ne' dalla madre: $(1-p(T|P)) \cdot (1-p(T|P)) = 0.25$

b) Il figlio di Tizio, avendo un genitore sano (S), potrà essere portatore sano (P) solo se riceve l'anomalia da Tizio, il che avviene solo se quest'ultimo è M oppure P. Come in altri esercizi di questo blocco, dobbiamo fare una sorta di media ponderata delle due probabilità di Trasmissione condizionate a M e a P, con pesi dati dalle probabilità appena calcolate che l'individuo di prima generazione Tizio sia M ovvero P:

$$p(T) = p(T|M) \cdot p(M) + p(T|P) \cdot p(P) = p(T|M) \cdot 0.25 + p(T|P) \cdot 0.50 = 1 \cdot 0.25 + 0.50 \cdot 0.50 = 0.50$$

Nota: nella somma sopra manca un pezzo relativo all'ultima possibilità per Tizio, che è essere sano. Questo pezzo è pari a zero, per questo la sua presenza non è stata esplicitata:

$$p(T|S) \cdot p(S) = 0 \cdot p(S)$$

ES 4 (Malformazioni in gravidanza)

Questo è il classico schema Binomiale: ci interessa l'evento ("successo") "trovare una malformazione", che ha probabilità $\pi=0.15$, osservato in $n=8$ "prove" (test in gravidanza). X =numero di successi nelle 8 prove.

Domande: $\Pr(X=2)$ e $\Pr(X=0)$

$$p(X=2) = \binom{8}{2} 0.15^2 \cdot 0.85^6 = \frac{8 \cdot 7 \cdot 6 \cdot \dots \cdot 2 \cdot 1}{(2 \cdot 1) \cdot (6 \cdot 5 \cdot \dots \cdot 2 \cdot 1)} 0.15^2 \cdot 0.85^6 = 0.237$$

$$p(X=0) = \binom{8}{0} 0.15^0 \cdot 0.85^8 = \frac{8!}{0!(8-0)!} 1 \cdot 0.85^8 = 0.272$$

ES 5 (Morti per influenza A H1N1)

Il dato è $p(\text{morte} \mid \text{malato}) = \pi = 0.0001$. Si vuole conoscere la probabilità che il numero di morti X fra $n=100$ individui malati sia pari a 0. Si tratterebbe di uno schema Binomiale, ma data la bassa probabilità dell'evento e l'alto numero di prove, si può applicare la formula della Poisson (che viene infatti detta anche legge degli "eventi rari").

Il parametro della Poisson è $\lambda = n \cdot \pi = 0.01$

$$p(X=0) = \frac{e^{-0.01} 0.01^0}{0!} = e^{-0.01} = 0.99$$

ES 6 (Peso alla nascita)

Questo è un semplice problema sull'utilizzo della curva Normale. I parametri sono: media $\mu = 2.8$, varianza $\sigma^2 = 0.16$

Quindi, deviazione standard $\sigma = 0.4$.

$p(X < 1.8)$ = probabilità sulla $N(0,1)$ che Z sia $<$ del valore standardizzato z :

$$z = \frac{1.8 - 2.8}{0.4} = -2.5$$

La tabella non ci fornisce $\Phi(z)$ (l'area fino a z) se z è negativo; ma questa area è uguale a quella nella coda destra, e cioè a $1 - \Phi(2.5)$;

sulle tavole in dotazione agli Studenti (che potrebbero essere diverse se prese da testi di statistica: attenzione!), in corrispondenza della riga 2.5 e della colonna 0, leggo $\Phi(2.5) = 0.994 \rightarrow$ la probabilità cercata è $0.006 \approx 6\%$.

ES 7 (Test diagnostico per la microcitemia)

In questo esercizio, abbiamo un test diagnostico che ha:

$p(P \mid M)$ = sensibilità = 80%

$p(N \mid \text{non } M)$ = specificità = 98% (N = test Negativo)

La "malattia" (essere microcitemico) ha $p(M) = 15\%$

Il numero di test positivi è dato dalla somma dei Veri Positivi (i Positivi per i Microcitemici) più i Falsi Positivi (i Positivi per i non-Microcitemici). I Microcitemici dovrebbero essere 15 (sempre il 15% dei bambini osservati!), gli altri 85 sono non-Microcitemici. Dunque:

- il numero dei Veri Positivi atteso è: $p(P \mid M) \cdot 15 = 0.8 \cdot 15 = 12$

- il numero dei Falsi Positivi atteso è: $p(P \mid \text{non } M) \cdot 85 = (1 - 0.98) \cdot 85 = 1.7$

$\rightarrow$ in totale: circa 14 test risulteranno Positivi.

ES 8 (screening allergie agli alimenti)

sensitività = 90% = (introduco un po' di notazione) $p(TP | AL)$

specificità = 60% = $p(N | \text{non } M)$

Nella seconda parte del quesito, abbiamo anche la prevalenza, diversa a seconda dell'età del soggetto: $Pr(AL)$ è 0.6 per i bambini e 0.015 per gli adulti

a) Indica troppi falsi negativi? **Falso**: i falsi negativi sono il 10% (1-sensitività)

b) Indica troppi falsi positivi? **Vero**: i falsi positivi sono il 40% (1-specificità)

c) Conduce spesso ad una diagnosi errata di allergia alimentare, mentre il soggetto non è allergico? **Vero**, è la stessa affermazione del punto b)

d) E' molto probabile che un soggetto allergico risulti negativo al test? **Falso**, è la stessa affermazione del punto a)

e) Un bambino sottoposto allo Skin-prick-test ha test positivo con probabilità 6%?

Falso. La probabilità di Test Positivo va calcolata nel seguente modo:

Sottintendiamo che parliamo di bambini. TP = "Test Positivo". AL = "Allergico".

$$\begin{aligned} Pr(TP) &= Pr((TP \& AL) \text{ oppure } (TP \& \text{non } AL)) = \\ &= Pr(TP \& AL) + Pr(TP \& \text{non } AL) = \\ &= Pr(TP | AL) \cdot Pr(AL) + Pr(TP | \text{non } AL) \cdot Pr(\text{non } AL) = \\ &= \text{sens.} \cdot 0.06 + (1-\text{spec}) \cdot (1 - 0.06) = \\ &= 0.9 \cdot 0.06 + 0.4 \cdot 0.94 = 0.43 \end{aligned}$$

Quindi 43% corregge 6%.

f) Un bambino sottoposto allo Skin-prick-test ha test positivo con probabilità 90%?

Falso: la correzione è "Un bambino allergico" (come suggerisce la seguente affermazione)

g) Un adulto allergico sottoposto allo Skin-prick-test ha test positivo con probabilità 90%? **Vero**

h) Un bambino con Skin-prick-test positivo ha probabilità di essere allergico pari a 13%?

Vero: risulta calcolando il valore predittivo del test positivo:

Sottintendiamo che parliamo di bambini. TP = "Test Positivo". AL = "Allergico".

$$\begin{aligned} Pr(AL | TP) &= \text{sens.} \cdot \text{prev.} / [\text{sens.} \cdot \text{prev.} + (1-\text{spec}) \cdot (1 - \text{prev})] = \\ &= 0.9 \cdot 0.06 / 0.9 \cdot 0.06 + 0.4 \cdot 0.94 = 0.126 \end{aligned}$$

i) La probabilità di essere allergico quando lo Skin-prick-test è positivo è la stessa per adulti e bambini?

Falso, poiché per gli adulti la prevalenza è più bassa, ossia è minore la probabilità a priori di essere allergico, e quindi lo sarà anche la probabilità a posteriori. Verifichiamo facendo il calcolo, in maniera analoga al punto precedente:

Stavolta parliamo di adulti. TP = "Test Positivo". AL = "Allergico".

$$\begin{aligned} Pr(AL | TP) &= \text{sens.} \cdot \text{prev.} / [\text{sens.} \cdot \text{prev.} + (1-\text{spec}) \cdot (1 - \text{prev})] = \\ &= 0.9 \cdot 0.015 / 0.9 \cdot 0.015 + 0.4 \cdot 0.985 = 0.033 \end{aligned}$$

Per gli adulti, la prob. di essere allergici avendo avuto un test positivo è solo del 3.3%.