14.4 - Falsi positivi, falsi negativi

TEST DIAGNOSTICI: FALSI POSITIVI, FALSI NEGATIVI

Nella vita purtroppo capita (ad alcuni più sovente che ad altri) di essere sottoposti ad esami medici: che siano del sangue, o radiologici, o di qualsivoglia tipo, essi generalmente sono finalizzati a verificare se si è affetti o meno da una data patologia.

Se risulto "POSITIVO" all'esame vuol dire che PROBABILMENTE SONO MALATO.
Per la maggior parte dei test clinici, questo "probabilmente" non vuol dire "sicuramente":
potrei infatti essere un "FALSO POSITIVO",
ossia potrebbe capitare che l'esame indichi erroneamente che sono malato mentre in realtà non lo sono.

Se risulto "NEGATIVO" all'esame vuol dire che PROBABILMENTE SONO SANO;
tuttavia, in genere, questo "probabilmente" non è una sicurezza completa;
potrei infatti essere un "FALSO NEGATIVO",
ossia potrebbe capitare che l'esame indichi erroneamente che sono sano mentre in realtà sono malato.

In medicina di solito si indaga sulla possibile presenza di una patologia mediante un test preliminare, poco costoso e/o poco "invasivo", per trarne una *prima indicazione* sulla probabilità che la malattia ci sia o non ci sia; eventualmente, se lo ritiene opportuno, il medico potrà poi prescrivere indagini più accurate e specialistiche, le quali dovrebbero stabilire con certezza pressoché assoluta la verità.

Tutto ciò si può riassumere nella tabella seguente:

		Esito del test preliminare	
		Positivo	Negativo
Condizione reale	Malati	Veri Positivi (VP)	Falsi Negativi (**FN**)
	Sani	Falsi Positivi (**FP**)	Veri Negativi (VN)

Un test diagnostico è "buono", è "affidabile",
nella misura in cui tende, in *presenza* di malattia, a fornire esito "positivo",
e nella misura in cui tende, in *assenza* di malattia, a fornire esito "negativo".

Questi due aspetti della bontà di un test
(rilevare effettivamente la malattia, se questa è presente, non rilevarla se questa è assente)
vengono chiamati rispettivamente la sua "sensibilità" e la sua "specificità".

$$\text{SENSIBILITA' di un test} = \text{probabilità che questo risulti positivo, nel caso si sia malati} = p(\mathrm{T}^+/\mathrm{M})$$

$$\text{SPECIFICITA' di un test} = \text{probabilità che questo risulti negativo, nel caso si sia sani} = p(\mathrm{T}^-/\mathrm{S})$$

La sensibilità e la specificità di un test si stabiliscono tramite studi molto approfonditi nell'ambito dei quali la presenza o assenza della malattia in ognuno dei soggetti partecipanti all'indagine vengono riconosciute con sicurezza perché a tal fine si fa uso di metodi *più generali* e/o *più sofisticati* rispetto al test stesso.

Le cosiddette ricerche "epidemiologiche", poi, permettono di determinare, in una data popolazione, la "PREVALENZA" della malattia, ossia la percentuale della popolazione che ne è portatrice, o, il che è lo stesso, la probabilità, per un soggetto "pescato" a caso nella popolazione, di essere malato.

$$\begin{aligned}&\text{PREVALENZA di una malattia in una popolazione} =\\&= \text{percentuale di malati sul totale della popolazione} =\\&= \text{probabilità, per un individuo preso a caso nella popolazione, di essere malato} =\\&= p(\mathrm{M}) = \frac{\textit{numero malati}}{\textit{numero totale di individui nella popolazione}}\end{aligned}$$

Si dice "VALORE PREDITTIVO POSITIVO" di un test la probabilità,
per un individuo che si è sottoposto al test e ha avuto esito positivo, di avere realmente la malattia:

$$\text{VALORE PREDITTIVO POSITIVO di un test (V.P.P.)} = \text{prob., se il test è positivo, di essere davvero malati}$$

$$\text{V.P.P.} = p(\mathrm{M}/\mathrm{T}^+) = \frac{n^\circ\ \textit{malati fra i positivi}}{n^\circ\ \textit{positivi}} \quad \textit{opp.} \quad \text{V.P.P.} = p(\mathrm{M}/\mathrm{T}^+) \overset{\textit{Bayes}}{=} \frac{p(\mathrm{M})\cdot p(\mathrm{T}^+/\mathrm{M})}{p(\mathrm{M})\cdot p(\mathrm{T}^+/\mathrm{M}) + p(\mathrm{S})\cdot p(\mathrm{T}^+/\mathrm{S})}$$

Si dice "VALORE PREDITTIVO NEGATIVO" di un test la probabilità,
per un individuo che si è sottoposto al test e ha avuto esito NEGATIVO, di essere realmente sano:

$$\text{VALORE PREDITTIVO NEGATIVO di un test (V.P.N.)} = \text{prob., se il test è negativo, di essere davvero sani}$$

$$\text{V.P.N.} = p(\mathrm{S}/\mathrm{T}^-) = \frac{n^\circ\ \textit{sani fra i negativi}}{n^\circ\ \textit{negativi}} \quad \textit{opp.} \quad \text{V.P.N.} = p(\mathrm{S}/\mathrm{T}^-) \overset{\textit{Bayes}}{=} \frac{p(\mathrm{S})\cdot p(\mathrm{T}^-/\mathrm{S})}{p(\mathrm{S})\cdot p(\mathrm{T}^-/\mathrm{S}) + p(\mathrm{M})\cdot p(\mathrm{T}^-/\mathrm{M})}$$

ESEMPIO 1 (PRELIMINARE: si fanno conteggi su un gran numero di persone,
poi da questi conteggi si traggono valutazioni di probabilità)

Da studi epidemiologici si sa che una data malattia è presente nel 4% della popolazione considerata.
Ora, un determinato test (ancora in fase sperimentale … pochissimo costoso, ma anche assai poco affidabile!)

♪ se praticato su di una persona effettivamente malata,
produce esito positivo nel 90% dei casi (Veri Positivi) e negativo nel rimanente 10% (Falsi Negativi)

♫ se praticato su di una persona sana,
produce esito negativo nel 95% dei casi (Veri Negativi) e positivo nel rimanente 5% (Falsi Positivi)

Se si prendono10000 persone a caso nella popolazione, e le si sottopone tutte e 10000 al test,

a) quanti pressappoco saranno davvero i Malati?
b) quanti pressappoco saranno davvero i Sani?
c) quanti saranno i Veri Positivi al test?
d) quanti saranno i Falsi Negativi al test?
e) quanti saranno i Falsi Positivi al test?
f) quanti saranno i Veri Negativi al test?
g) quanti saranno *in totale* i Positivi al test?
h) quanti saranno *in totale* i Negativi al test?
i) per un Malato, qual è la probab. di essere Positivo?
j) per un Malato, qual è la probab. di essere Negativo?
k) per un Sano, qual è la probab. di essere Negativo?
l) per un Sano, qual è la probab. di essere Positivo?
m) per un Positivo, qual è la probab. di essere Malato?
n) per un Negativo, qual è la probab. di essere Malato?
o) per un Negativo, qual è la probab. di essere Sano?
p) per un Positivo, qual è la probab. di essere Sano?

RISPOSTE:

a) $\mathrm{M} = 10000 \cdot 4/100 = 400$

b) $\mathrm{S} = 10000 - 400 = 9600$

c) $\mathrm{VP} = 400 \cdot \frac{90}{100} = 360$

d) $\mathrm{FN} = 400 \cdot \frac{10}{100} = 40 \;\; (\textit{oppure}: 400 - 360 = 40)$

e) $\mathrm{FP} = 9600 \cdot \frac{5}{100} = 480$

f) $\mathrm{VN} = 9600 \cdot \frac{95}{100} = 9120$

g) $\mathrm{P} = \mathrm{VP} + \mathrm{FP} = 360 + 480 = 840$

h) $\mathrm{N} = \mathrm{VN} + \mathrm{FN} = 9120 + 40 = 9160$

i) $p(\mathrm{T}^+ / \mathrm{M}) = \frac{\mathrm{VP}}{\mathrm{M}} = \frac{360}{400} = 0{,}90$ (*già lo si sapeva*)

j) $p(\mathrm{T}^- / \mathrm{M}) = 1 - 0{,}90 = 0{,}10$ (*già lo si sapeva*)

k) $p(\mathrm{T}^- / \mathrm{S}) = \frac{\mathrm{VN}}{\mathrm{S}} = \frac{9120}{9600} = 0{,}95$ (*già lo si sapeva*)

l) $p(\mathrm{T}^+ / \mathrm{S}) = 1 - 0{,}95 = 0{,}05$ (*già lo si sapeva*)

m) $\underbrace{p(\mathrm{M} / \mathrm{T}^+)}_{\text{V.P.P.}} = \frac{\mathrm{VP}}{\mathrm{P}} = \frac{360}{840} \approx 0{,}4286$

n) $p(\mathrm{M} / \mathrm{T}^-) = \frac{\mathrm{FN}}{\mathrm{N}} = \frac{40}{9160} \approx 0{,}0044$

o) $\underbrace{p(\mathrm{S} / \mathrm{T}^-)}_{\text{V.P.N.}} = \frac{\mathrm{VN}}{\mathrm{N}} = \frac{9120}{9160} \approx 0{,}9956$

p) $p(\mathrm{S} / \mathrm{T}^+) = \frac{\mathrm{FP}}{\mathrm{P}} = \frac{480}{840} \approx 0{,}5714$

OSSERVAZIONE
Le probabilità di cui ai punti m), n), o), p) si sarebbero potute calcolare anche applicando il **Teorema di Bayes**:

m) $\underbrace{p(\mathrm{M}/\mathrm{T}^+)}_{\text{V.P.P.}} = \frac{p(\mathrm{M}) \cdot p(\mathrm{T}^+/\mathrm{M})}{p(\mathrm{M}) \cdot p(\mathrm{T}^+/\mathrm{M}) + p(\mathrm{S}) \cdot p(\mathrm{T}^+/\mathrm{S})} = \frac{0{,}04 \cdot 0{,}90}{0{,}04 \cdot 0{,}90 + 0{,}96 \cdot 0{,}05} = \frac{0{,}036}{0{,}036 + 0{,}048} = \frac{0{,}036}{0{,}084} \approx 0{,}4286$

n) $p(\mathrm{M}/\mathrm{T}^-) = \frac{p(\mathrm{M}) \cdot p(\mathrm{T}^-/\mathrm{M})}{p(\mathrm{M}) \cdot p(\mathrm{T}^-/\mathrm{M}) + p(\mathrm{S}) \cdot p(\mathrm{T}^-/\mathrm{S})} = \frac{0{,}04 \cdot 0{,}10}{0{,}04 \cdot 0{,}10 + 0{,}96 \cdot 0{,}95} = \frac{0{,}004}{0{,}004 + 0{,}912} = \frac{0{,}004}{0{,}916} \approx 0{,}0044$

o) $\underbrace{p(\mathrm{S}/\mathrm{T}^-)}_{\text{V.P.N.}} = \frac{p(\mathrm{S}) \cdot p(\mathrm{T}^-/\mathrm{S})}{p(\mathrm{S}) \cdot p(\mathrm{T}^-/\mathrm{S}) + p(\mathrm{M}) \cdot p(\mathrm{T}^-/\mathrm{M})} = \frac{0{,}96 \cdot 0{,}95}{0{,}96 \cdot 0{,}95 + 0{,}04 \cdot 0{,}10} = \frac{0{,}912}{0{,}912 + 0{,}004} = \frac{0{,}912}{0{,}916} \approx 0{,}9956$

m) $p(\mathrm{S}/\mathrm{T}^+) = \frac{p(\mathrm{S}) \cdot p(\mathrm{T}^+/\mathrm{S})}{p(\mathrm{S}) \cdot p(\mathrm{T}^+/\mathrm{S}) + p(\mathrm{M}) \cdot p(\mathrm{T}^+/\mathrm{M})} = \frac{0{,}96 \cdot 0{,}05}{0{,}96 \cdot 0{,}05 + 0{,}04 \cdot 0{,}90} = \frac{0{,}048}{0{,}048 + 0{,}036} = \frac{0{,}048}{0{,}084} \approx 0{,}5714$

ESERCIZI

1) Una patologia virale infetta il 3% della popolazione di un dato territorio. Il test diagnostico più diffuso, su 100 persone sane che si sottopongono al test, mediamente diagnostica per errore 2 falsi positivi; e in compenso su 100 malati che si sottopongono al test, mediamente diagnostica per errore 2 falsi negativi.
   a) Se una persona ha avuto esito positivo al test, che probabilità c'è che sia davvero malata?
   b) E se ha avuto esito negativo al test, che probabilità c'è che sia davvero immune?
   c) Se la percentuale dei malati nella popolazione raddoppiasse, come cambierebbero tali due probabilità?

2) Un test per la gravidanza è ancora in fase sperimentale.
Negli studi per perfezionarlo, fino ad ora si sono sottoposte al test 912 donne, delle quali la metà davvero incinte e la metà no, coi risultati raccolti nella tabella qui a fianco.

	incinte	non incinte
positive	424	10
negative	32	446

Si domanda: se il test venisse commercializzato così com'è,
a) una donna davvero incinta, che probabilità avrebbe, pressappoco, di risultare positiva?
b) E una donna che risultasse positiva, che probabilità avrebbe di essere davvero incinta?
c) Se una donna risultasse negativa, che probabilità avrebbe di essere invece incinta?

3) Imposta un foglio elettronico in modo che l'utente possa inserire le probabilità relative ad un problema sui test diagnostici e gli venga calcolata la risposta.
Utilizzalo poi per fare simulazioni, variando prevalenza, sensibilità e specificità.

prevalenza=p(M)=	0,08	p(M/T+)=VPP	0,892018779
da cui p(S)	0,92	p(S/T-)=VPN	0,99562746
p(T+/M)=sensibilità=	0,95	**Gli unici dati che l'utente deve inserire**	
p(T+/S), conseguenza della specificità (qui sotto)	0,01	**sono quelli nelle celle grigie:**	
p(T-/S)=specificità=	0,99	**tutti gli altri valori li deve calcolare**	
p(T-/M), conseguenza della sensibilità (in alto)	0,05	**automaticamente il foglio elettronico**	

4) Supponiamo che, per un test diagnostico, la probabilità di risultare positivi se malati sia del 96% e la probabilità di risultare negativi se si è sani sia del 98%.
Supponiamo di somministrare il test a tappeto in una popolazione di 10000 individui.
Allora il numero di falsi positivi e di falsi negativi dipende dalla prevalenza della malattia nella popolazione, ossia dalla percentuale di popolazione affetta dalla malattia.
Calcola il numero atteso di falsi positivi e di falsi negativi sotto l'ipotesi
a) che la prevalenza della patologia sia del 5% b) che la prevalenza della patologia sia dell'1%.

5)

	T –	T +	
M –	2280	5	2285
M+	27	124	151
	2307	129	2436

La tabella qui a fianco è stata riempita sperimentando un nuovo test diagnostico su 2436 persone (i veri malati sono stati riconosciuti attraverso indagini più costose e accurate).
Quanti falsi positivi ci sono stati? Quanti falsi negativi?
In che percentuale (arrotondando all'intero) si può valutare il V.P.P. (= Valore Predittivo Positivo) di questo test? E in che percentuale (sempre arrotondando all'intero) il V.P.N. (Valore Predittivo Negativo)?

6) Una accurata ricerca clinica ha mostrato che un determinato virus è presente nel 2% della popolazione.
Il test più diffuso per rilevare la presenza del virus (pericoloso soltanto qualora la persona abbia il sistema immunitario compromesso) ha una "sensibilità" del 97% e una "specificità" del 999 per 1000.
Calcolare il "V.P.P." (= Valore Predittivo Positivo) del test.

7) Se la specificità di un test è 1, cosa si può dire del Valore Predittivo Positivo di quel test?

Il seguente problema, preso dal sito www.ufl.edu della University of Florida, ha un risultato che forse non ci si aspetta, ma di cui, riflettendo, si comprende bene la ragione!

8) Per un test diagnostico, sono uguali al 95% tanto la sensibilità quanto la specificità.
La prevalenza della malattia nella popolazione è dell'1%.
Una persona viene selezionata a caso nella popolazione, e sottoposta al test.
Se putacaso questa persona risulta positiva, che probabilità ha di essere davvero malata?

RISPOSTE

1) a) $p(\mathrm{M}/\mathrm{T}^+)=\dfrac{p(\mathrm{M})\cdot p(\mathrm{T}^+/\mathrm{M})}{p(\mathrm{M})\cdot p(\mathrm{T}^+/\mathrm{M})+p(\mathrm{S})\cdot p(\mathrm{T}^+/\mathrm{S})}=\dfrac{0{,}03\cdot 0{,}98}{0{,}03\cdot 0{,}98+0{,}97\cdot 0{,}02}\approx 60\%$

b) $p(\mathrm{S}/\mathrm{T}^-)=\dfrac{p(\mathrm{S})\cdot p(\mathrm{T}^-/\mathrm{S})}{p(\mathrm{S})\cdot p(\mathrm{T}^-/\mathrm{S})+p(\mathrm{M})\cdot p(\mathrm{T}^-/\mathrm{M})}=\dfrac{0{,}97\cdot 0{,}98}{0{,}97\cdot 0{,}98+0{,}03\cdot 0{,}02}\approx 99{,}9\%$ (*il calcolo dà* 0,999369...)

c) Se la percentuale dei malati nella popolazione raddoppiasse, la 1ª percentuale diventerebbe del 76% circa e la 2ª, pur diminuendo leggerissimamente, sarebbe ancora vicina al 99,9% (il calcolo dà 0,998699…)

2) Arrotondando all'unità: a) ≈ 93% b) ≈98% c) ≈ 7% 4) a) 190; 20 b) 198; 4 5) 5; 27; 96%; 99%

6) $\text{V.P.P.}=p(\mathrm{V}/\mathrm{T}^+)=\dfrac{p(\mathrm{V})\cdot p(\mathrm{T}^+/\mathrm{V})}{p(\mathrm{V})\cdot p(\mathrm{T}^+/\mathrm{V})+p(\overline{\mathrm{V}})\cdot p(\mathrm{T}^+/\overline{\mathrm{V}})}=\dfrac{0{,}02\cdot 0{,}97}{0{,}02\cdot 0{,}97+0{,}98\cdot 0{,}001}\approx 95\%$ 7) che è anch'esso =1

8) ≈ 16%. Perché così bassa? Logico: i sani nella popolazione sono tanti, e il 5% dei sani, se testato, risulta falsamente positivo, quindi se la persona - estratta a caso fra tutta la popolazione - riporta dal test esito positivo, è più facile che si tratti di un falso positivo che di un malato … Meno male!