1. Teorema di Bayes

Il teorema di Bayes può essere applicato in medicina nei test diagnostici. Questi test non sono purtroppo infallibili e a volte un individuo sano può risultare positivo al test (si dice in questo caso che è un *falso positivo*) oppure un individuo malato può risultare negativo al test (si dice in questo caso che è un *falso negativo*). Nella conduzione di un test diagnostico è necessario conoscere a priori:

- il valore di **specificità**, cioè la probabilità che su un paziente sano il test risulti negativo
- il valore di **sensibilità**, cioè la probabilità che su un paziente che è malato il test risulti positivo.

I valori di probabilità coinvolti in un test diagnostico si possono schematizzare nella seguente tabella:

Soggetto sottoposto al test		
Esito del Test	**Sano**	**Malato**
Negativo	$p(T^- \mid S)$	$p(T^- \mid M)$
Positivo	$p(T^+ \mid S)$	$p(T^+ \mid M)$

dove:

- $p(T^- \mid S)$ indica la specificità (paziente sano con test negativo)
- $p(T^- \mid M)$ indica la probabilità di avere un paziente malato con test negativo (falso negativo)
- $p(T^+ \mid S)$ indica la probabilità di avere un paziente sano con test positivo (falso positivo)
- $p(T^+ \mid M)$ indica la sensibilità (paziente malato con test positivo).

Per il significato che abbiamo attribuito a queste probabilità appare evidente che:

$$p(T^- \mid S) + p(T^+ \mid S) = 1 \qquad \text{e} \qquad p(T^- \mid M) + p(T^+ \mid M) = 1$$

Normalmente i valori $p(T^- \mid S)$ e $p(T^+ \mid M)$ sono noti, in quanto ricavati da un periodo di sperimentazione del test; di conseguenza, in base alle relazioni precedenti, si possono ritenere noti anche i valori $p(T^- \mid M)$ e $p(T^+ \mid S)$.

Le questioni importanti di fronte all'esito di un test diagnostico sono le seguenti:

- qual è la probabilità $p(M \mid T^+)$ che un soggetto positivo al test sia effettivamente malato
- qual è la probabilità $p(S \mid T^-)$ che un soggetto negativo al test sia effettivamente sano.

Per rispondere a queste due domande si può usare il teorema di Bayes.

Nel primo caso: $$p(M \mid T^+) = \frac{p(T^+ \mid M) \cdot p(M)}{p(T^+ \mid S) \cdot p(S) + p(T^+ \mid M) \cdot p(M)}$$

Nel secondo: $$p(S \mid T^-) = \frac{p(T^- \mid S) \cdot p(S)}{p(T^- \mid S) \cdot p(S) + p(T^- \mid M) \cdot p(M)}$$

Per poter applicare queste due formule è necessario conoscere rispettivamente $p(M)$, cioè la probabilità di essere effettivamente malati, e $p(S)$, cioè la probabilità di essere sani. Di solito questi valori di probabilità derivano dai dati statistici sulla diffusione della patologia in questione e sono quindi valori noti.
Nel foglio di Excel a pagina seguente abbiamo calcolato i due valori di probabilità $p(M \mid T^+)$ e $p(S \mid T^-)$ nel caso di una malattia che, da rilevazioni statistiche, colpisce il 4% della popolazione, con un valore di specificità uguale al 92,8% e un valore di sensibilità del 98,4%.

Nella tabella in alto a sinistra abbiamo inserito i due valori di specificità e di sensibilità e calcolato per differenza con 1 gli altri due valori della colonna:

B4: = 1 – B3

C3: = 1 – C4

Nella cella B6 abbiamo inserito il valore $p(M) = 0{,}04$ e in quella sottostante abbiamo calcolato per differenza $p(S)$:

B7: = 1 – B6

Siamo pronti per calcolare quanto richiesto nelle celle F2 e F4:

F2: = (C4*B6) / (B4*B7 + C4*B6) probabilità che un soggetto positivo al test sia effettivamente malato

F4: = (B3*B7) / (B3*B7 + C3*B6) probabilità che un soggetto negativo al test sia effettivamente sano

	A	B	C	D	E	F
1		Soggetto				
2	Test	Sano	Malato		$p(M\|T^+) =$	0,362832
3	Negativo	0,928	0,016			
4	Positivo	0,072	0,984		$p(S\|T^-) =$	0,999282
5						
6	$p(M) =$	0,04				
7	$p(S) =$	0,96				

Dai valori calcolati possiamo fare qualche considerazione sul test in questione:

- poiché $p(M \mid T^+) = 0{,}362832$, è abbastanza bassa la probabilità che il soggetto sia effettivamente malato una volta che il test sia risultato positivo
- poiché $p(S \mid T,^-) = 0{,}999282$, è molto alta la probabilità che il soggetto sia sano se il test è negativo.