

Metodologia clinica - giovanni.boniolo@unife.it

GIOVANNI BONIOLO

**Dipartimento di Scienze Biomediche e Chirurgico Specialistiche
Università of Ferrara
Via Fossato di Mortara, 64A
44121-Ferrara**

Corso breve di metodologia clinica

Ore	Giorno	Argomento
2	Mercoledì 8/11- 9,00-11,00	Introduzione - Sensibilità e specificità dei test clinici. Falso positivo e negativo
4	Lunedì 13/11 – 11,00-13,00	Sensibilità e specificità dei test clinici –Approccio Bayesiano – Valore predittivo positivo e negativo
6	Lunedì 20/11 -11,00-13,00	Curva ROC
8	Mercoledì 22/11 – 9,00-11,00	1. Introduction 2. Confidence Interval of a Proportion 3. The Standard Deviation 4. The Gaussian Distribution
10	Lunedì 4/12 – 11,00-13,00	5. The Confidence Interval of a Mean 10. What is a P Value? 11. Statistical Significance and Hypothesis Testing 12. Interpreting Significant and Not Significant P Value
12	Mercoledì 6/12 – 9,00-11,00	Out of order
14	Lunedì 11/12 – 11,00-13,00	14. Interpreting Lab Tests: Introduction to Bayesian Thinking 15. Bayes and Statistical Significance 16. Bayes Theorem in Genetics
16	Mercoledì 13/12 – 9,00-11,00	
18	Lunedì 18/12 – 11,00	Esame scritto con idoneità
20	Mercoledì 20/12 – 9,00-11,00	
24	Lunedì 8/01 – 11,00	Esame scritto con idoneità

Testi consigliati

H.C.Sox, M.C. Higgins, D.K. Owens, Medical decision making, Blackwell

H. Motulsky, Intuitive biostatistics, Oxford University Press

Testi sito docente

H. Motulsky, Intuitive biostatistics, Oxford University Press (old edition),
Ch.1,2,3,4,5,10,11,12,14,15,16

T.W. Loong Understanding sensitivity and specificity with the right side of the brain. BMJ.
2003 Sep 27;327(7417):716-9

T.Fawcett An introduction to ROC analysis, Pattern Recognition Letters 27 (2006) 861–874

R. Nuzzo, Scientific method: statistical errors. Nature. 2014 Feb 13;506(7487):150-2

V. Velickovic, What Everyone Should Know about Statistical Correlation. American Scientist,
19/02/16

Augusto Murri, *Lezioni di clinica medica*, Società Editrice Libreria, Milano, 1908

Enrico Poli, *Metodologia medica: principi di logica e pratica clinica*, Rizzoli Editore, 1965

Cesare Scandellari, *La strategia della diagnosi*, Piccin Editore, Padova 1981

Mario Austoni, Giovanni Federspil, *Principi di Metodologia Clinica*, Cedam ed., Padova, 1975

Giovanni Federspil, *I fondamenti del metodo scientifico in medicina clinica e sperimentale*.
Piccin Editore. Padova 1980

Giovanni Federspil, *Logica Clinica. I princìpi del metodo in medicina*, Mc Graw-Hill, Milano
2004.

Cesare Scadellari, *La diagnosi clinica*, Masson, Milano 2005

Metodologia medica o clinica: insieme delle regole che guidano l'agire e il ragionare sul malato

Sintomo e segno

Valore segnaletico del segno: capacità di indicare una possibile malattia

Valore probativo del segno: capacità di provare la presenza di una malattia

Sindrome: raggruppamento di segni caratteristicamente associati a una malattia e codificati come tali dalla nosografia medica

Complesso sindromico: raggruppamento di segni realizzato dal medico relativamente al quel dato paziente durante il procedimento diagnostico

Normalità e patologia

Normale rispetto a un dato ambiente, epoca.

Frequente perché normale non normale perché frequente

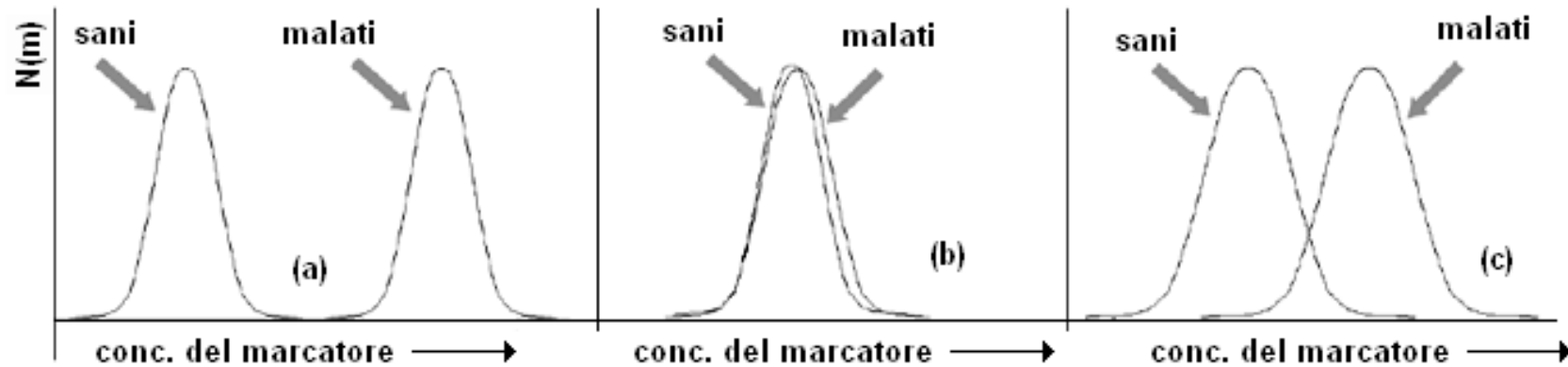
Rilievi clinici a due valori (presenza-assenza) e a più valori discreti o continui. Necessità in tal caso di trovare un valore di soglia

Gold standard: valore di un rilievo clinico con il quale si indica la presenza o l'assenza sicura di una patologia

Raramente un valore di soglia è un gold standard. Di solito da un test di laboratorio che non fornisce un gold standard comporta che vi sia un intervallo di valori che possono essere trovati sia nella popolazione portatrice di quella malattia, sia nella popolazione sana o portatrice di un'altra malattia

Di solito si realizza una sovrapposizione come quella sotto

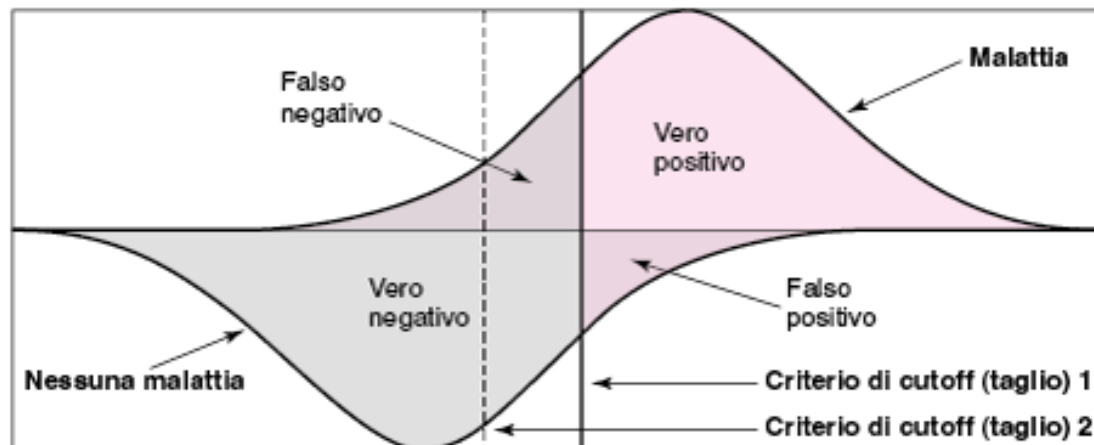
L'importanza della sensibilità di un test diagnostico

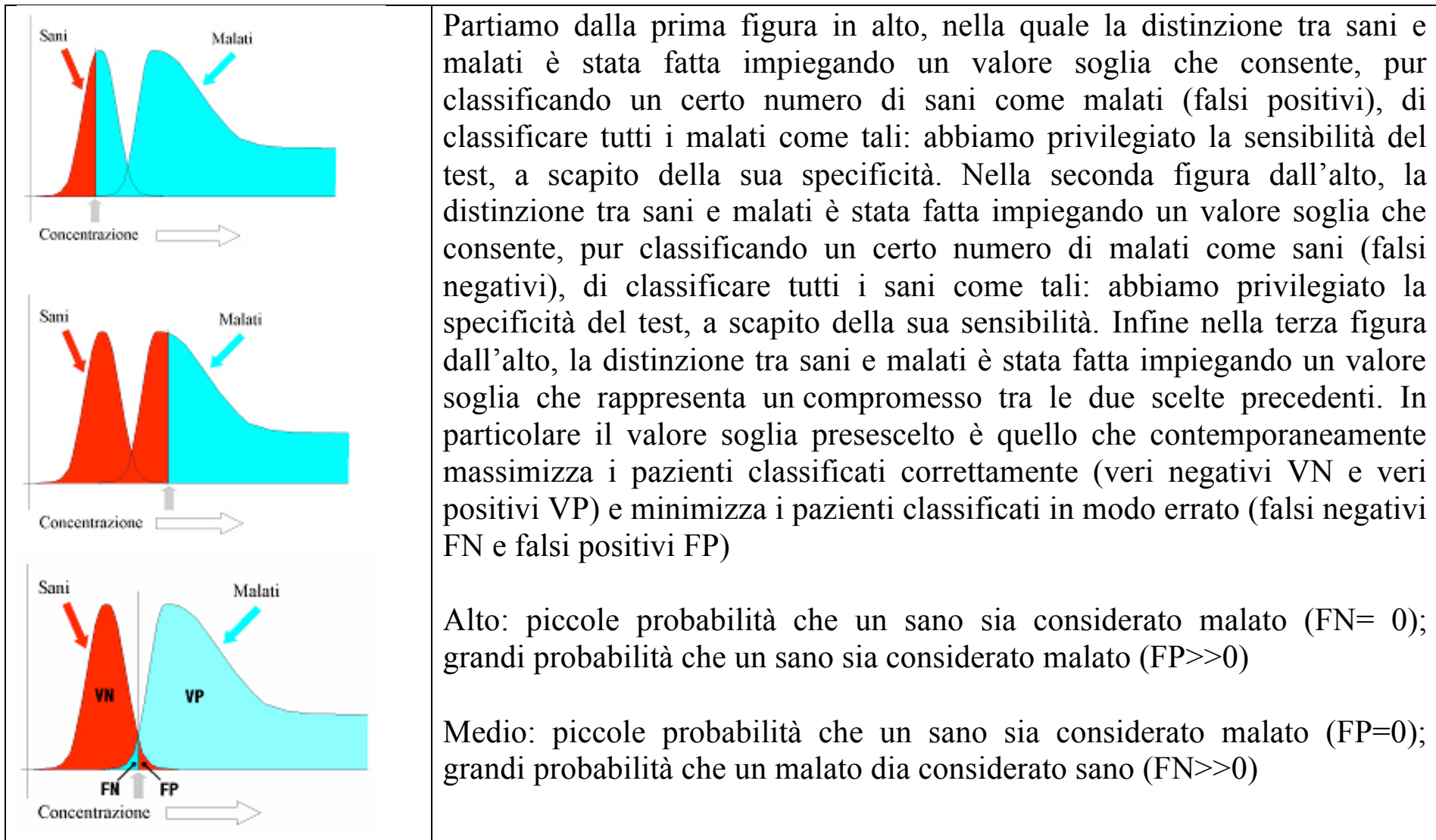


I tre grafici riportano in ordinate il numero di persone, $N(m)$, in cui è presente una certa concentrazione di marcatore tumorale, m ; in ascisse sono indicate le concentrazioni del marcatore. Il grafico (a) schematizza il test ideale in cui la differenza di concentrazione del marcatore è netta e permette di distinguere tra malato e non malato; il grafico (b) si riferisce ad un test inutile perché il marcatore è inadatto; il grafico (c) si riferisce alla tipica situazione reale dove c'è una zona di sovrapposizione fra malato e non malato.

Distribuzioni dei risultati del test.

I pazienti con la malattia sono mostrati nella distribuzione superiore; i pazienti sani sono mostrati nella distribuzione inferiore. Per i pazienti con malattia, la regione sotto la distribuzione dei risultati che si trova a destra (sopra) rispetto al valore del punto limite corrisponde alla percentuale dei veri positivi (ossia la sua sensibilità); l'area che si trova al di sotto a sinistra del valore limite corrisponde alla percentuale di falsi-negativi. Per i pazienti sani, la regione a destra del valore del punto limite corrisponde alla percentuale di falsi positivi, e la regione a sinistra corrisponde invece alla percentuale di veri negativi (ossia la sua specificità). Per le 2 parti sovrapposte delle distribuzioni (p.es., pazienti con e senza malattia), la variazione della linea del valore limite influenza la sensibilità e la specificità, ma in direzioni opposte; cambiando il valore limite da 1 a 2 si riduce il numero di falsi negativi (maggiore sensibilità), ma aumenta anche il numero di falsi positivi (minore specificità).





Partiamo dalla prima figura in alto, nella quale la distinzione tra sani e malati è stata fatta impiegando un valore soglia che consente, pur classificando un certo numero di sani come malati (falsi positivi), di classificare tutti i malati come tali: abbiamo privilegiato la sensibilità del test, a scapito della sua specificità. Nella seconda figura dall'alto, la distinzione tra sani e malati è stata fatta impiegando un valore soglia che consente, pur classificando un certo numero di malati come sani (falsi negativi), di classificare tutti i sani come tali: abbiamo privilegiato la specificità del test, a scapito della sua sensibilità. Infine nella terza figura dall'alto, la distinzione tra sani e malati è stata fatta impiegando un valore soglia che rappresenta un compromesso tra le due scelte precedenti. In particolare il valore soglia prescelto è quello che contemporaneamente massimizza i pazienti classificati correttamente (veri negativi VN e veri positivi VP) e minimizza i pazienti classificati in modo errato (falsi negativi FN e falsi positivi FP)

Alto: piccole probabilità che un sano sia considerato malato ($FN=0$); grandi probabilità che un sano sia considerato malato ($FP \gg 0$)

Medio: piccole probabilità che un sano sia considerato malato ($FP=0$); grandi probabilità che un malato dia considerato sano ($FN \gg 0$)

Falso positivo

Da Wikipedia, l'enciclopedia libera.

In [statistica](#) il **falso positivo** è il risultato di un [test](#) che porta erroneamente a rifiutare l'[ipotesi](#) sulla quale esso è stato condotto.

Più in generale, in qualunque ambito in cui si presenti una decisione predittiva binaria (positivo o negativo), un **falso positivo** indica la scelta *a torto* di "positivo", ovvero un *falso allarme*. Un esempio in [informatica](#) è un [antivirus](#) che considera erroneamente dannoso un programma innocuo.

In [statistica](#) questo errore del test è detto **errore di primo tipo**. L'altro possibile errore è quello di *secondo tipo*, che genera [falsi negativi](#).

Ipotesi	accettata	rifiutata
valida	buona decisione vero positivo	errore di primo tipo: falso positivo
sbagliata	errore di secondo tipo: falso negativo	buona decisione vero negativo

Falso negativo

Da Wikipedia, l'enciclopedia libera.

In [statistica](#) il **falso negativo** è il risultato di un [test](#) che porta erroneamente ad accettare l'[ipotesi](#) sulla quale esso è stato condotto.

Più in generale, in qualunque ambito in cui si presenti una decisione predittiva binaria (positivo o negativo), un falso negativo indica la scelta "a torto" di "negativo". Un esempio in [informatica](#) è un filtro [antispam](#) che lasci erroneamente passare una lettera indesiderata.

In [statistica](#) questo errore del test è detto errore di secondo tipo. L'altro possibile errore è quello di primo tipo, che genera [falsi positivi](#).

ipotesi	accettata	rifiutata
valida	buona decisione vero positivo	errore di primo tipo: falso positivo
sbagliata	errore di secondo tipo: falso negativo	buona decisione vero negativo

Rilievo patognomonico: se positivo è condizione sufficiente per la malattia ma se negativo non sufficiente per la sua assenza. Comporta zero FP

Rilievo obbligatorio: se negativo è condizione sufficiente per l'assenza della malattia ma se positivo non sufficiente per la sua presenza. Comporta zero FN

Rilievo perfetto: comporta zero FN e zero FP (gold standard)

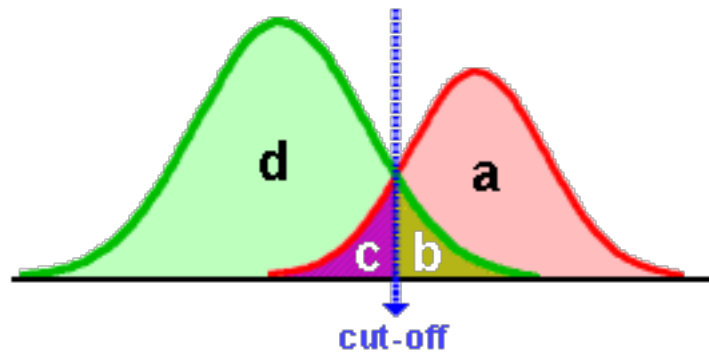
Supponiamo che una persona si rechi da un medico e questi decida di prescrivergli un test per verificare la presenza di una sospetta situazione patologica. Il paziente potrebbe chiedere: “Se il risultato del test fosse positivo, significherebbe che sono malato?”. Per poter rispondere, il medico dovrebbe conoscere e aver compreso alcuni elementi di statistica. In particolare,

- i) la tabulazione della popolazione rispetto a tale test e tale malattia;
- ii) il teorema di Bayes e quanto da esso deriva.

Prima di tutto, ogni test è caratterizzato da una data *sensibilità* e da una data *specificità*, dove:

- **la sensibilità** è la probabilità che il test presenti esito positivo in presenza di malattia;
- **la specificità** è la probabilità che il test presenti esito negativo in assenza di malattia.

Sensibilità e specificità sono probabilità dei patologi



	M+	M-
T+	a	b
T-	c	d

$$Se = a / (a+c)$$

$$Sp = d / (d+b)$$

Sia N il numero di individui (la popolazione) su cui il test viene effettuato e che ha sviluppato (o meno) una malattia pienamente diagnosticata. Possiamo riassumere i risultati nella tavola sottostante:

	M	S	
T ⁺	a (VP)	b (FP)	a+b (VP+FP)
T ⁻	c (FN)	d (VN)	c+d (FN+VN)
	a+c M=(VP+FN)	b+d S=(FP+VN)	N= a+b+c+d

M = numero individui affetti dalla malattia;

S = numero individui sani;

T⁺ = numero individui positivi al test;

T⁻ = numero individui negativi al test;

a, b, c, d = numero degli individui caratterizzati da quella condizione (M or S) e dal risultato del test (T⁺ or T⁻).

Con ciò abbiamo che:

$$\text{Sensibilità} = P(T^+/M) = \frac{a}{a+c} = \frac{VP}{VP+FN} = \frac{VP}{M}$$

$$\text{Specificità} = P(T^-/S) = \frac{d}{b+d} = \frac{VN}{FP+VN} = \frac{VN}{S}$$

Naturalmente il test potrebbe non essere perfetto e di conseguenza potrebbe esserci

- *frazione di risultati falsi-positivi* (il test è positivo, ma non c'è la malattia)
- *frazione di risultati falsi-negativi* (il test è negativo, ma la malattia è presente). Dove

$$\text{Frazione falsi-positivi} = P(T^+/S) = \frac{b}{b+d} = 1 - \text{specificità}$$

$$\text{Frazione falsi-negativi} = P(T^-/M) = \frac{c}{a+c} = 1 - \text{sensibilità}$$

Pertanto un medico che prescrive un test dovrebbe conoscerne sensibilità e specificità (e, di conseguenza, la frazione di falsi-positivi e di falsi-negativi correlate). Già a questo punto, il medico potrebbe rispondere a una domanda che potrebbe venire sollevata dall'ipotetico paziente seduto di fronte a lui: **“Qual è la probabilità che io abbia realmente la malattia se il test risulta positivo?”** oppure **“Qual è la probabilità che io non abbia realmente la malattia se il test risulta negativo?”**.

In altre parole, il medico potrebbe calcolare la probabilità di avere la malattia nonostante il risultato positivo del test (*valore predittivo positivo*), o la probabilità di non avere la malattia nonostante il risultato negativo del test (*valore predittivo negativo*).

$$\text{valore predittivo positivo} = P(M/T^+) = \frac{a}{a+b} = \frac{VP}{VP+FP}$$

$$\text{valore predittivo negativo} = P(S/T^-) = \frac{d}{c+d} = \frac{VN}{FN+VN}$$

In effetti, queste probabilità risultano significative soltanto se riferite alla **prevalenza** della presunta malattia nella popolazione di interesse, aspetto, quest'ultimo, che costituisce un ulteriore dato che il medico dovrebbe conoscere. Maggiore è la diffusione della malattia nella popolazione considerata, più alto sarà il valore predittivo positivo (e il numero di casi positivi scoperti). Da notare che mentre la sensibilità e la specificità sono caratteristiche del test, i valori predittivi positivo e negativo dipendono dalla prevalenza della malattia nella popolazione considerata.

Valore predittivo negativo e positivo sono probabilità del clinico

C'è, tuttavia, un modo migliore di mettere insieme le osservazioni raccolte (sul test e sulla popolazione considerata) e la correlazione fra il valore predittivo positivo/negativo e la prevalenza della malattia nella popolazione considerata. Questa modalità è offerta dal **teorema di Bayes**.

Incidenza = (Numero di nuovi ammalati in un dato periodo / popolazione) K

Prevalenza = (Numero di ammalati / popolazione) K = Incidenza x Durata malattia = $\frac{M}{M+S} = \frac{VP+FN}{M+S}$

K = costante (1000 o 100.000) che aumenta la leggibilità

Esempio: incidenza morbillo nel 2017 = 13 (casi morbillo) / 168.000 (popolazione osservata) = 0,000077

Se K = 100.000 allora Incidenza morbillo = 7,7 (rischio di avere morbillo ogni 100.000 casi)

Casi reali : Popolazione = Incidenza : K

Non risulta rilevante in questo contesto comprendere in che modo il teorema di Bayes sia stato formulato e dimostrato, quanto invece sapere che esso permette al medico di collegare i dati che riguardano la sensibilità e la specificità del test con la diffusione della malattia in un modo esplicito. Più precisamente, esso si configura quale strumento formale che permette di passare dal livello patologico che riguarda la conoscenza degli effetti date la cause (riassunto in $P(T^+/M)$) al livello clinico che riguarda la conoscenza delle cause dati gli effetti (riassunto in $P(M/T^+)$). Il teorema di Bayes dice che

$$P(H / E) = \frac{P(E / H)P(H)}{P(E)}$$

Dove

- $P(H/E)$, è la *probabilità a posteriori* ed è la probabilità condizionata dell'ipotesi H data l'evidenza E; ossia è il grado di credenza che H sia vero tenendo in considerazione E;
- $P(E/H)$ è la probabilità condizionata di E data H;
- $P(H)$, è la *probabilità a priori* di H, ossia è il grado di credenza iniziale in in H.
- $P(E)$ è la probabilità di E.

Nel nostro caso, la probabilità di avere la malattia M (ipotesi) se il test T (l'evidenza) è positivo – ovvero, $P(M/T^+)$ - e considerata prevalenza della malattia nella popolazione di interesse (ovvero, il valore predittivo positivo), diviene:

$$P(M / T^+) = \frac{P(T^+ / M)P(M)}{P(T^+)} = \frac{\text{sensibilità} \cdot \text{prevalenza}}{(a + b) / N}$$

dove, come detto, sensibilità = $P(T^+/M)$ e prevalenza = $P(M)$ ¹.

Risulta abbastanza intuitivo comprendere perché $P(M/T^+)$ è una probabilità *a posteriori* (è la probabilità della malattia una volta effettuato il test) e $P(M)$ è la probabilità *a priori* (è la probabilità della malattia prima di aver effettuato il test). Risulta importante considerare che, poiché la sensibilità e la prevalenza sono al numeratore della frazione, maggiori sono (come fattori indipendenti), più alto sarà il loro valore positivo predittivo.

¹ Ovviamente, se $P(M) = \text{prevalenza}$, allora $P(S) = 1 - P(M) = 1 - \text{prevalenza}$.

Non difficilmente, possiamo anche definire:

1) il valore predittivo negativo, ovvero la probabilità di avere un risultato negativo al test senza avere la patologia

$$P(S/T^+) = \frac{P(T^+ / S)P(S)}{P(T^+)} = \frac{\text{specificità} \cdot (1 - \text{prevalenza})}{(c + d) / N}$$

2) la probabilità di un falso positivo al test senza avere la malattia:

$$P(T^+ / S) = \frac{P(S / T^+)P(T^+)}{P(S)} = \frac{(1 - \text{sensibilità}) \cdot (a + b) / N}{1 - \text{prevalenza}}$$

Illustriamo quanto sopra considerando il caso di un test che scopre un marker per il tumore al colon retto, cioè il M2-PK².

Supponiamo che sia un test molto affidabile, con una sensibilità dell'85% (cioè 0,85), e una specificità del 95% (cioè 0,95).

Sappiamo anche che il rischio cumulativo di tumore al colon retto in individui con un'età inferiore ai 75 anni di età è del 3,9% (cioè 0,039) in Europa.

A questo punto noi possiamo costruire la tabella di contingenza (si veda sotto) considerando una popolazione di 100.000 individui e calcolando sia i risultati predittivi positivi che quelli negativi.

² È la forma dimerica del piruvato chinasi di tipo M2.

	M = individui con il cancro al colon retto (confermato tramite colonscopia)	S = individui non affetti da cancro al colon retto (confermato tramite colonscopia)	
T ⁺ = M2-PK il test è positivo	2,797 (a= veri-positivi)	4,835 (b = falsi-positivi)	7,632
T ⁻ = M2-PK il test è negativo	493 (c = falsi-negativi)	91,875 (d= veri-negativi)	92,368
	3,290	96,710	N=100.000

Per cui si ha il **valore predittivo positivo**

$$P(M / T^+) = \frac{\text{sensibilità} \cdot \text{prevalenza}}{(a+b) / N} = \frac{0,85 \cdot 0,039}{\frac{7632}{100000}} = 0,43 \quad (43\%)$$

e il **valore predittivo negativo**

$$P(S / T^+) = \frac{\text{specificità} \cdot (1 - \text{prevalenza})}{(c+d) / N} = \frac{0,95 \cdot (1 - 0,039)}{\frac{92368}{100000}} = 0,99 \quad (99\%)$$

Ora possiamo domandare: “ Il test è clinicamente rilevante?”.

Noi sappiamo che ciò dipende dalla prevalenza, poiché dipende dalla differenza fra il dato relativo all'incidenza della malattia sulla popolazione considerata prima e dopo il test.

Nel nostro caso, la probabilità, prima del test, di avere un tumore al colon retto era del 3,9%.

Dopo il test sappiamo che, se il risultato del test è positivo, noi abbiamo una probabilità del 43% di sviluppare la malattia e, se il test è negativo, del 99% di non svilupparla.

Ciò significa che effettuare il test porta a un reale incremento di conoscenza e che quindi ha senso farlo.

Fallacia della probabilità primaria

Si considera un'informazione irrilevante o comunque non la si contestualizza alla popolazione in questione

Esempio

Test con Sensibilità=0,8=80%; Specificità=0,9=90%; Malattia con prevalenza pari a 1%

Tavola di contingenza su una popolazione di 100.000 persone:

	M	S	
T ⁺	800 (VP)	9.900 (FP)	10.700 (VP+FP)
T ⁻	200 (FN)	89.100 (VN)	89.300 (FN+VN)
	1.000 M=(VP+FN)	99.000 S=(FP+VN)	N= a+b+c+d= 100.000

Se il test ha una sensibilità dell'80% e siamo positivi, erroneamente potremmo pensare di avere la patologia all'80%. **Ma questa è una fallacia** perché non si è tenuto conto della prevalenza, cioè della malattia rispetto alla popolazione

Infatti se la prevalenza è dell'1%, allora

$$P(M/T^+) = 0,074 = 7,4\%$$

Per cui, per 100.000 persone, 1000 sono malate e 99.000 sono sane. Se si fa un test ci sono 9.900 sani che risultano negativi e 10.700 che sani che risultano positivi, mentre ci sono 800 malati che risultano positivi. Per cui se siamo malati siamo uno degli 800 sui 10700, ovvero $800/10.700 = 0,074 = 7,4\%$

Con Bayes

$$P(M/T^+) = \frac{0,8 \times 0,01}{\frac{10.700}{100.000}} = \frac{0,8 \times 0,01}{0,107} = 0,074$$

Dimostrazione Teorema di Bayes

1)

$$\text{Sensibilità} = P(T^+/M) = \frac{VP}{VP+FN} = \frac{VP}{M}$$

$$\text{Prevalenza} = P(M) = \frac{M}{M+S} = \frac{VP+FN}{M+S}$$

$$\text{A: Sensibilità} \times \text{Prevalenza} = P(T^+/M) \times P(M) = \frac{VP}{M} \frac{M}{M+S} = \frac{VP}{M+S}$$

2)

$$\text{valore predittivo positivo} = P(M/T^+) = \frac{VP}{VP+FP}$$

$$\text{probabilità di avere un test positivo} = P(T^+) = \frac{VP+FP}{M+S}$$

$$\text{B: valore predittivo positivo} \times \text{probabilità di avere un test positivo} = P(M/T^+) \times P(T^+) = \frac{VP}{VP+FP} \frac{VP+FP}{M+S} = \frac{VP}{M+S}$$

Poiché A=B

$$P(T^+/M) \times P(M) = P(M/T^+) \times P(T^+)$$

Ossia

$$P(M/T^+) = P(T^+/M) \times P(M) / P(T^+) \text{ c.v.d.}$$

Dove

$P(M/T^+)$ probabilità a posteriori

$P(T^+/M)$ verisimiglianza (likelihood)

$P(M)$ probabilità a priori

Dal punto di vista epistemologico risultano evidenti nel teorema di Bayes:

- la capacità di formalizzare il meccanismo con cui si ribalta la logica, da *l'effetto|data la causa a la causa|dato l'effetto*, passando dalla deduzione all'induzione (che risulta per definizione solo probabile);
- la capacità di formalizzare il meccanismo con cui l'informazione fornita dall'esperienza (lo specifico risultato del test di laboratorio) si somma all'informazione a priori, aumentando la nostra conoscenza;
- la capacità di *misurare l'informazione* che un test diagnostico fornisce alla diagnosi medica

Dal punto di vista pratico risulta evidente dal teorema di Bayes che:

- in condizioni di bassa prevalenza diminuisce il valore predittivo del test positivo;
- in condizioni di bassa specificità del test diminuisce il valore predittivo del test positivo;
- in condizioni di bassa specificità e di bassa prevalenza aumenta il valore predittivo del test negativo, quindi un test diventa utile soprattutto per escludere la malattia.

		predicted condition			
		prediction positive	prediction negative		
total population				Prevalence = $\frac{\Sigma \text{condition positive}}{\Sigma \text{total population}}$	
true condition	condition positive	True Positive (TP)	False Negative (FN) (type II error)	True Positive Rate (TPR), Sensitivity, Recall, Probability of Detection = $\frac{\Sigma \text{TP}}{\Sigma \text{condition positive}}$	False Negative Rate (FNR), Miss Rate = $\frac{\Sigma \text{FN}}{\Sigma \text{condition positive}}$
	condition negative	False Positive (FP) (Type I error)	True Negative (TN)	False Positive Rate (FPR), Fall-out, Probability of False Alarm = $\frac{\Sigma \text{FP}}{\Sigma \text{condition negative}}$	True Negative Rate (TNR), Specificity (SPC) = $\frac{\Sigma \text{TN}}{\Sigma \text{condition negative}}$
Accuracy = $\frac{\Sigma \text{TP} + \Sigma \text{TN}}{\Sigma \text{total population}}$		Positive Predictive Value (PPV), Precision = $\frac{\Sigma \text{TP}}{\Sigma \text{prediction positive}}$	False Omission Rate (FOR) = $\frac{\Sigma \text{FN}}{\Sigma \text{prediction negative}}$	Positive Likelihood Ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic Odds Ratio (DOR) = $\frac{\text{LR}+}{\text{LR}-}$
		False Discovery Rate (FDR) = $\frac{\Sigma \text{FP}}{\Sigma \text{prediction positive}}$	Negative Predictive Value (NPV) = $\frac{\Sigma \text{TN}}{\Sigma \text{prediction negative}}$	Negative Likelihood Ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

Curva ROC

ROC è l'acronimo di *Receiver Operating Characteristic* (caratteristiche operative del ricevitore). Il ricevitore in questione è il radar.

Durante la seconda guerra mondiale il problema era, nell'ambito del rumore di fondo presente nei segnali radar, riconoscere il segnale causato dalla presenza di oggetti nemici sui campi di battaglia (tipicamente nei cieli, ma anche in mare). Si tratta di un classico problema di rapporto segnale/rumore che è stato affrontato dagli ingegneri nell'ambito della teoria della rilevazione del segnale.

Vediamo un primo esempio numerico, rappresentato dalla seguente serie di dati:

3, 4, 6, 2, 6, 5, 4, 5, 3, 3, 6, 12, 4, 5, 6, 6, 7, 4, 5, 4, 4, 4, 3, 5, 11, 3, 3, 5, 2, 5, 4, 3, 2

Supponiamo di avere un filtro che possa essere variato in continuo, ponendo la prima volta la soglia a 12, e considerando "rumore" i valori inferiori a 12 e "segnale" i valori uguali o superiori a 12, ponendo la seconda volta la soglia a 11, e considerando "rumore" i valori inferiori a 11 e "segnale" i valori uguali o superiori a 11, e così via.

Dopo avere filtrato il segnale utilizzando tutte le possibili soglie (gli interi compresi tra 12 e 2) otteniamo la seguente serie di dati:

→ → → → → → → → → → → 12, →
→ → → → → → → → → → → 12, →
→ → → → → → → → → → → 12, →
→ → → → → → → → → → → 12, →
→ → → → → → → → → → → 12, →
→ → → → → → → → → → → 12, →
→ → 6, → 6, → → → → 6, 12, → → 6, 6, 7, → → → → → → → → → 11, → → → → → → → → →
→ → 6, → 6, 5, → 5, → → 6, 12, → 5, 6, 6, 7, → 5, → → → → 5, 11, → → 5, → 5, → → →
→ 4, 6, → 6, 5, 4, 5, → → 6, 12, 4, 5, 6, 6, 7, 4, 5, 4, 4, 4, → 5, 11, → → 5, → 5, 4, → →
3, 4, 6, → 6, 5, 4, 5, 3, 3, 6, 12, 4, 5, 6, 6, 7, 4, 5, 4, 4, 4, 3, 5, 11, 3, 3, 5, → 5, 4, 3, →
3, 4, 6, 2, 6, 5, 4, 5, 3, 3, 6, 12, 4, 5, 6, 6, 7, 4, 5, 4, 4, 4, 3, 5, 11, 3, 3, 5, 2, 5, 4, 3, 2

Supponendo che i segnali di intensità 11 e 12 provengano dagli oggetti che volevamo rilevare (quindi, nell'esempio aeronautico, da due aerei), abbiamo individuato ciò che ci interessava in mezzo al rumore di fondo del radar (valori da 7 in giù).

Il “rapporto segnale/rumore” (11 o 12 di segnale, contro 7 o meno di rumore) è elevato: e questo rende “evidente” la soglia tra rumore e segnale.

Ma si può avere a che fare con distribuzioni dei valori del tipo

3, 4, 6, 2, 6, 9, 5, 4, 10, 8, 10, 8, 5, 9, 6, 12, 11, 4, 7, 6, 8, 9, 3, 3, 5, 9, 12, 10, 4, 6, 7, 8, 6, 5, 4

Dopo avere filtrato il segnale utilizzando tutte le possibili soglie otteniamo

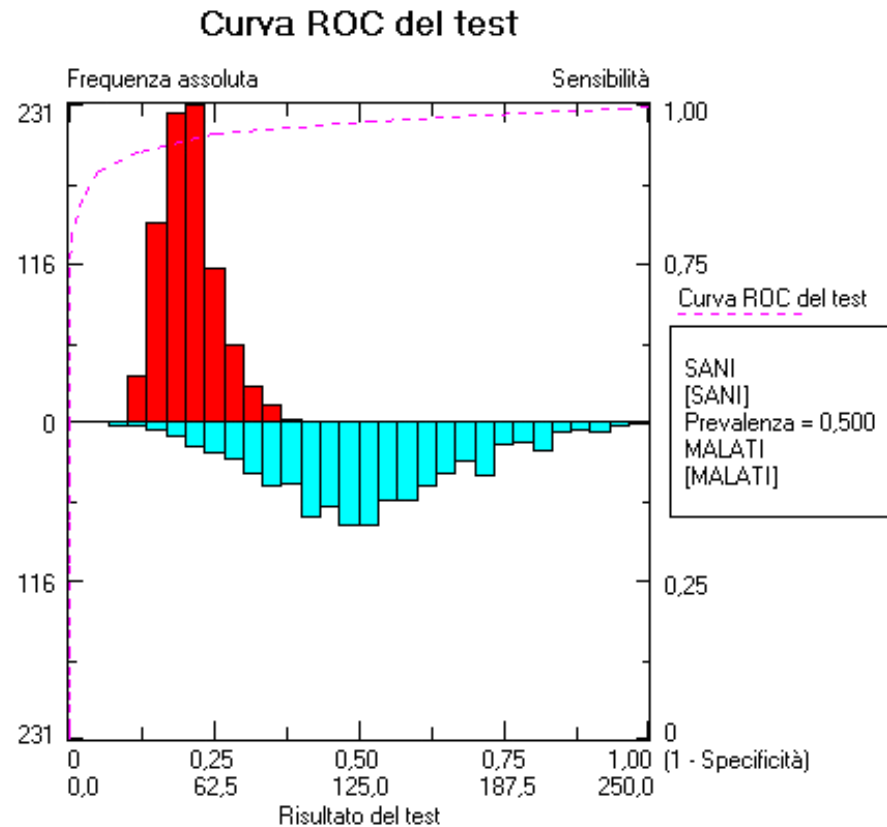
→ → → → → → → → → → → → → → → 12, → → → → → → → → → → → → → → → 12, → → → → → → → → →
→ → → → → → → → → → → → → → → 12, 11, → → → → → → → → → → → 12, → → → → → → → → →
→ → → → → → → → → 10, → 10, → → → → → 12, 11, → → → → → → → → → → → 12, 10, → → → → → → → → →
→ → → → → → 9, → → 10, → 10, → → 9, → 12, 11, → → → → → 9, → → → → 9, 12, 10, → → → → → → → → →
→ → → → → → 9, → → 10, 8, 10, 8, → 9, → 12, 11, → → → → 8, 9, → → → → 9, 12, 10, → → → → 8, → → → →
→ → → → → → 9, → → 10, 8, 10, 8, → 9, → 12, 11, → 7, → 8, 9, → → → → 9, 12, 10, → → → 7, 8, → → → →
→ → 6, → 6, 9, → → 10, 8, 10, 8, → 9, 6, 12, 11, → 7, 6, 8, 9, → → → → 9, 12, 10, → 6, 7, 8, 6, → → → →
→ → 6, → 6, 9, 5, → 10, 8, 10, 8, 5, 9, 6, 12, 11, → 7, 6, 8, 9, → → → 5, 9, 12, 10, → 6, 7, 8, 6, 5, →
→ 4, 6, → 6, 9, 5, 4, 10, 8, 10, 8, 5, 9, 6, 12, 11, 4, 7, 6, 8, 9, → → → 5, 9, 12, 10, 4, 6, 7, 8, 6, 5, 4
3, 4, 6, → 6, 9, 5, 4, 10, 8, 10, 8, 5, 9, 6, 12, 11, 4, 7, 6, 8, 9, 3, 3, 5, 9, 12, 10, 4, 6, 7, 8, 6, 5, 4
3, 4, 6, 2, 6, 9, 5, 4, 10, 8, 10, 8, 5, 9, 6, 12, 11, 4, 7, 6, 8, 9, 3, 3, 5, 9, 12, 10, 4, 6, 7, 8, 6, 5, 4

Il problema è che questa volta i valori variano da 2 a 12 senza soluzioni di continuo: e la soglia tra segnale e rumore non è più “evidente” come sembrava esserlo nel primo caso.

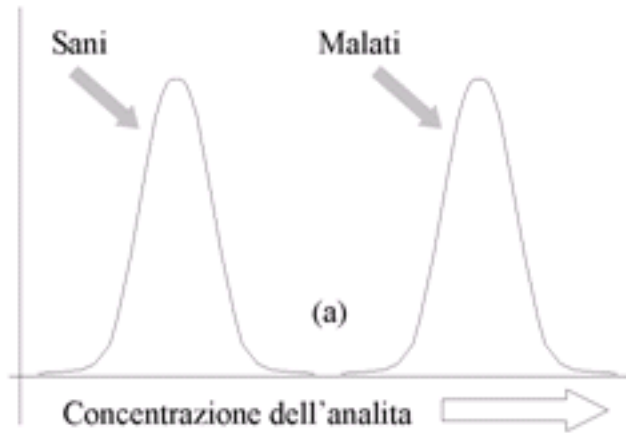
Cioè, qual è il valore che segna la transizione da un segnale a un non-segnale, ovvero quale è il valore soglia al di sotto del quale dobbiamo pensare che non si tratti più di un segnale (un aereo), ma si tratti semplicemente di rumore di fondo?

Questo secondo esempio illustra una situazione in cui il rapporto segnale/rumore è ridotto, e quindi abbiamo difficoltà a distinguere l'uno dall'altro. Gli ingegneri hanno modellizzato il problema trovando una generalizzazione che consente di affrontarlo. Lo stesso modello è utilizzato per la diagnostica di laboratorio e la diagnosi medica in generale. Nelle quali il risultato del test nei malati è il segnale, il risultato del test nei sani è il rumore (di fondo).

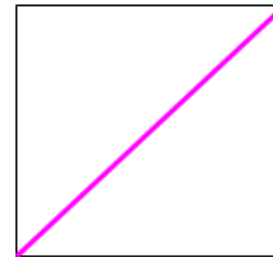
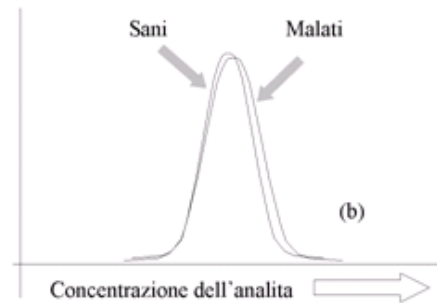
E la curva ROC è costruita semplicemente riportando in ascisse la sensibilità e in ordinate (1 - specificità)



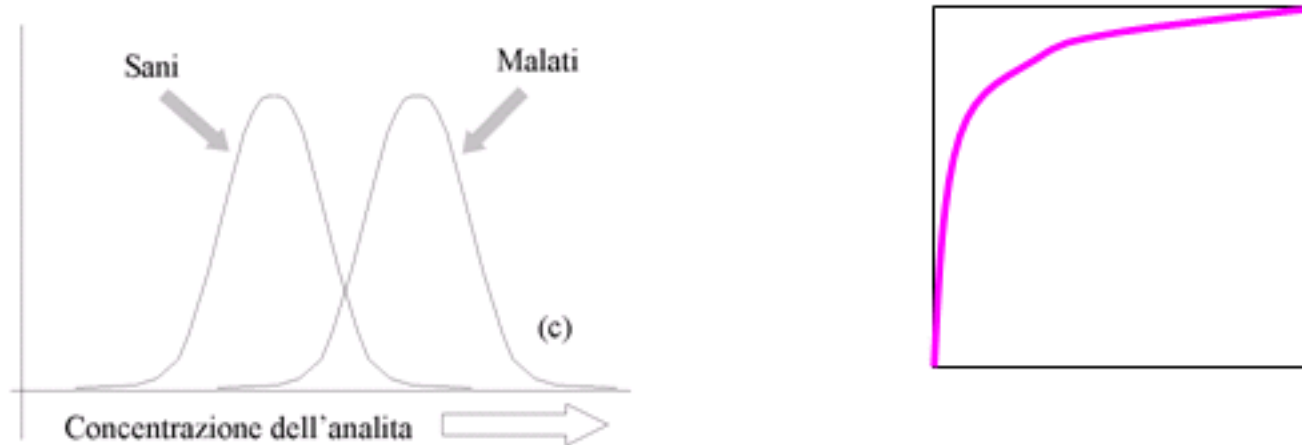
Nella situazione ideale, di completa separazione dei valori del rumore da quelli del segnale abbiamo una curva ROC che sale perfettamente verticale sull'asse delle ordinate quindi piega ad angolo retto in orizzontale, parallela all'asse delle ascisse:



Nella situazione opposta, di completa sovrapposizione dei valori del rumore e di quelli del segnale, nella quale segnale e rumore sono quindi indistinguibili l'uno dall'altro, abbiamo una curva ROC che è una retta che va dall'angolo inferiore sinistro all'angolo superiore destro:



Nella pratica si incontrano curve che giacciono in mezzo ai due estremi sopra riportati:



L'area sottesa alla curva fornisce una misura delle prestazioni del radar (nell'esempio aeronautico) e del test di laboratorio (nel nostro caso).

Il massimo valore di sensibilità è uguale a 1, e il massimo valore di $(1 - \text{specificità})$ è uguale a 1, e pertanto l'area totale massima sottesa dalla curva è uguale a 1.

Nel caso del radar l'area sottesa dalla curva rappresenta la probabilità che un aereo fornisca un segnale (radar) superiore a quello del rumore (di fondo). Nel caso del test diagnostico questa area rappresenta la probabilità che una persona con la malattia, presa a caso, abbia un risultato superiore a quello di una persona senza la malattia, presa a caso.

Se la curva ROC (del radar/del test di laboratorio) va dall'angolo inferiore sinistro all'angolo superiore destro la probabilità è uguale a 0,5: e l'informazione fornita dal radar/test di laboratorio è uguale a quella che si può ricavare dal lancio di una moneta!

L'unico limite delle curve ROC deriva dal fatto che sono costruite a partire dalla sensibilità e dalla specificità, e non prendono in considerazione la probabilità a priori (prevalenza della malattia nel caso di un test diagnostico).

Le curve ROC forniscono quindi una soluzione parziale, e per questo motivo non possono sostituire l'approccio bayesiano.

Per contro le curve ROC forniscono un facile e immediato strumento per il confronto delle prestazioni di un radar/test di laboratorio nei confronti di un altro.

Un radar le cui caratteristiche operative sono descritte da una curva che giace completamente al di sopra delle curve di un altro radar ha prestazioni migliori.

Analogamente un test di laboratorio descritto da una curva che giace completamente al di sopra delle curve di un altro test ha prestazioni migliori.

In entrambi i casi l'area sottesa dalla curva è una misura dell'informazione che il radar/test di laboratorio è in grado di fornire: tanto maggiore è l'area, tanto maggiore è l'informazione fornita (e tanto migliori sono il radar e il test di laboratorio).

Due sono i risultati notevoli.

Il primo è rappresentato dal fatto che la stessa soluzione è stata trovata partendo da problemi così diversi quali l'interpretazione di un segnale radar e l'interpretazione del risultato di un test di laboratorio.

Il secondo è rappresentato dal fatto che è possibile misurare l'informazione fornita da un test diagnostico sintetizzandola in un unico valore: cosa che rende possibile confrontare tra di loro le accuratezze diagnostiche di test differenti per la stessa malattia

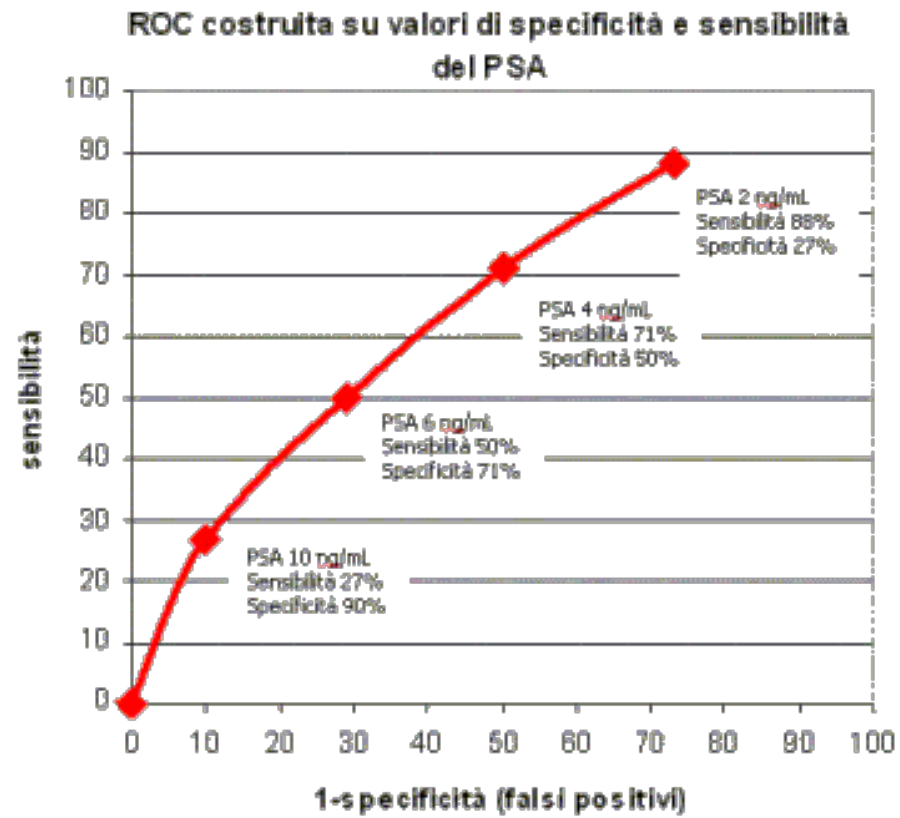
Si prenda per esempio il PSA, usato per lo screening del cancro della prostata.

Se si sceglie un cut-off di 4 ng/mL questo valore non garantisce di suddividere con un taglio netto i sani dai malati: vi saranno dei sani con valori superiori a 4 ng/mL (falsi positivi) e dei malati con valori inferiori (falsi negativi).

Se si decide di porre il cut-off ad un valore più basso (per esempio 2 ng/mL) sicuramente si avranno meno falsi negativi ma aumenteranno i falsi positivi.

Al contrario se si pone il cut-off ad un valore più elevato la maggiore specificità sarà scontata da una riduzione della sensibilità.

La rappresentazione grafica di tutto questo si può fare con un sistema di assi cartesiani in cui sull'asse delle ordinate si pone la sensibilità e su quello delle ascisse il numero dei falsi positivi (vale a dire $1 - \text{specificità}$). Per ogni valore di PSA si avranno valori di sensibilità e di specificità diversi e in questo modo ogni valore di PSA sarà individuato da un punto derivante dalla intersezione della sensibilità e della specificità rispettive. Unendo i vari punti così determinati si costruisce la **curva ROC**.



Per PSA di 10 ng/ml la sensibilità è bassa (cioè si perdono molti tumori) ma la specificità è elevata (cioè vi sono pochi falsi positivi).

Progressivamente aumenta la sensibilità e si riduce la specificità man mano che si abbassa il valore. Per valori di PSA di 2 ng/ml la sensibilità è massima (si identificano quasi tutti i casi di tumore) ma nello stesso tempo si avranno molti falsi positivi perché si riduce la specificità.

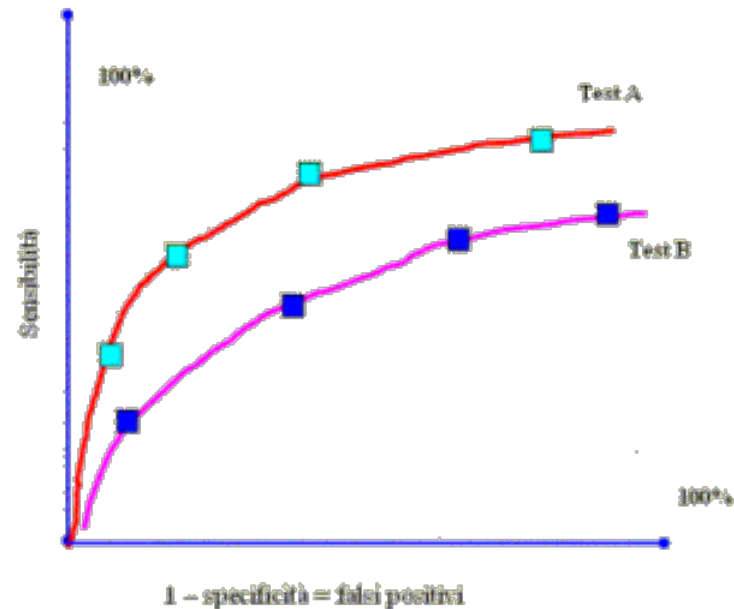
Nel decidere il cut-off di un esame conviene spesso scegliere un compromesso tra specificità e sensibilità, per esempio prendendo il punto della curva che più si avvicina all'angolo superiore sinistro del diagramma, in questo caso un valore compreso tra 4 e 6 ng/mL.

Si osservi che tanto più la curva si avvicina all'angolo superiore sinistro del diagramma tanto più ampia è l'area che essa sottende (area sotto la curva) e quindi tanto più efficace è il test.

Comunque non sempre, nella scelta del cut-off, si sceglie il punto che rappresenta il miglior compromesso tra sensibilità e specificità. Dipende anche dalla condizione che il test diagnostica.

Per esempio nel caso del dosaggio delle troponine, che servono a discriminare se vi è o meno una necrosi miocardica, si può privilegiare la sensibilità a scapito della specificità e quindi scegliere un cut-off basso (parte in alto a destra della curva) che consente di avere pochi falsi negativi pur potendosi avere un maggior numero di falsi positivi.

Le curve ROC permettono anche di paragonare l'accuratezza di due test usati per la diagnosi di una determinata malattia. Si supponga di avere due test per la diagnosi della malattia "X", il TEST A e il TEST B, le cui rispettive curve ROC sono esemplificate nella figura 2.



Come si può vedere il test A si avvicina di più all'angolo superiore sinistro del diagramma e la relativa curva sottotende un'area maggiore rispetto a quella sottesa dal test B. Il test A quindi avrà una performance superiore al test B.

Da notare infine che non necessariamente le curve ROC si costruiscono per test espressi con valori quantitativi, ma si possono disegnare anche per test che si esprimono in modo qualitativo. Si prenda per esempio la radiografia del torace per la diagnosi di polmonite. Si possono prevedere vari cut-off decisionali per il trattamento del paziente. Per esempio se il quadro radiologico viene considerato dal radiologo come probabilità “molto elevata di polmonite” e si decide di trattare solo questi casi si avrà una elevata specificità ma nello stesso tempo si possono perdere tutti i casi di polmonite che si esprimono con un quadro radiologico non patognomnico. Se invece si decide di trattare, oltre ai casi definiti certi e probabili, anche quelli descritti come “probabilità bassa di polmonite” si avrà una specificità ridotta (cioè si corre il rischio di trattare anche quadri che non sono una polmonite) ma con una sensibilità elevata (si tratteranno praticamente tutte le polmoniti). In un punto intermedio della curva si situeranno i casi definiti come probabilità intermedia di polmonite, come mostrato dalla figura.

