

CDL in MEDICINA & CHIRURGIA

Statistica Medica

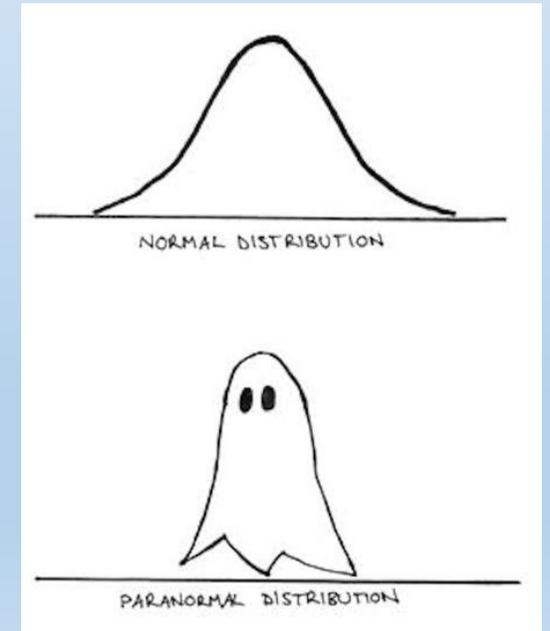
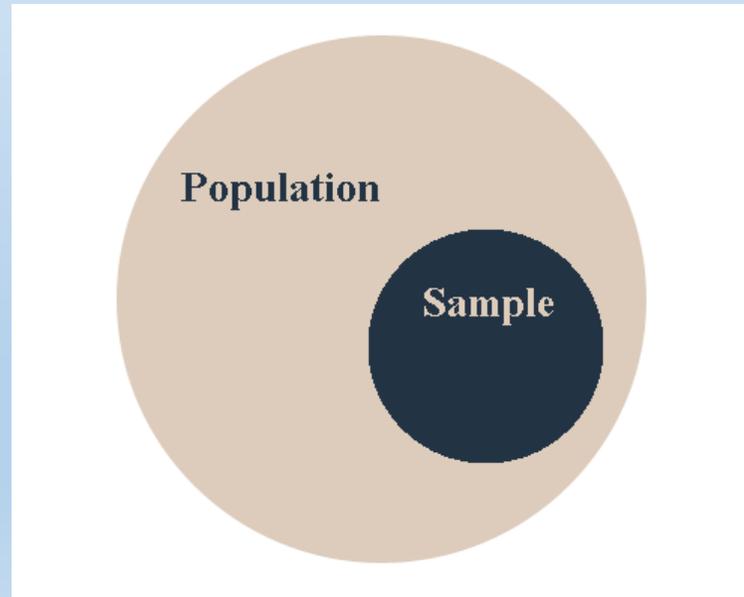
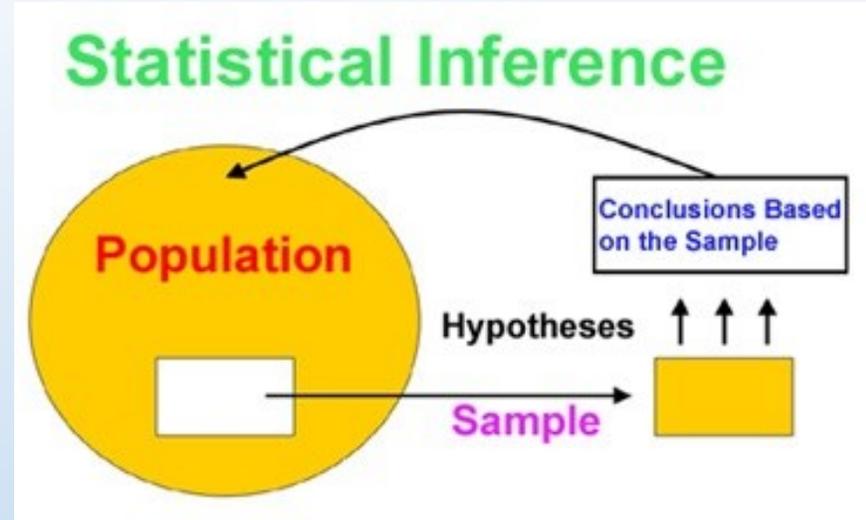
gbarbati@units.it

A.A. 2024-25



Sommario:

- Probabilità **condizionata** & Teorema di Bayes
- Test diagnostici
- Curva ROC



Probabilità condizionata

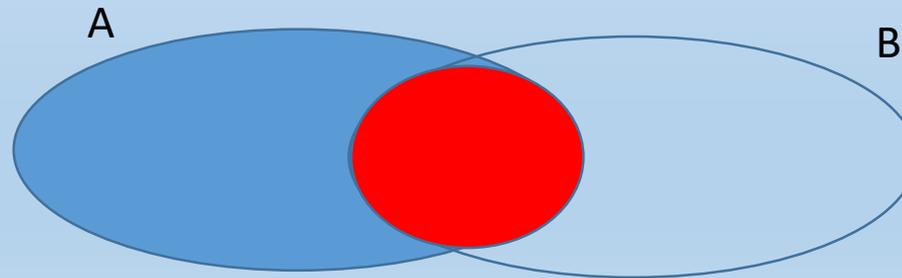
Se sappiamo che un evento si è verificato, come ciò **influisce** sulla probabilità che si sia verificato un altro evento?

Per rispondere a questa domanda, dobbiamo esaminare la probabilità **condizionata**.

Qual è la probabilità di sviluppare una malattia cardiovascolare (B), sapendo che si è diabetici (A)?

Qual è la probabilità di sviluppare un cancro al polmone (B), sapendo che si è fumatori (A)?

Qual è la probabilità di avere un ictus (B), sapendo che si soffre di ipertensione (A)?



Si considerino gli eventi:

X = estrazione di una **fiala X**

d = estrazione di una **fiala difettosa**

Qual è la probabilità di estrarre una fiala difettosa **dato che** la fiala è prodotta da X ?

FIALE	MACCHINE			X o Y o Z
	X	Y	Z	
\bar{d}	920	2760	5820	9500
d	80	240	180	500
d o \bar{d}	1000	3000	6000	10000

10000 fiale

Difettose (d)

Non difettose (\bar{d})

prodotte da 3 macchine (X, Y, Z).

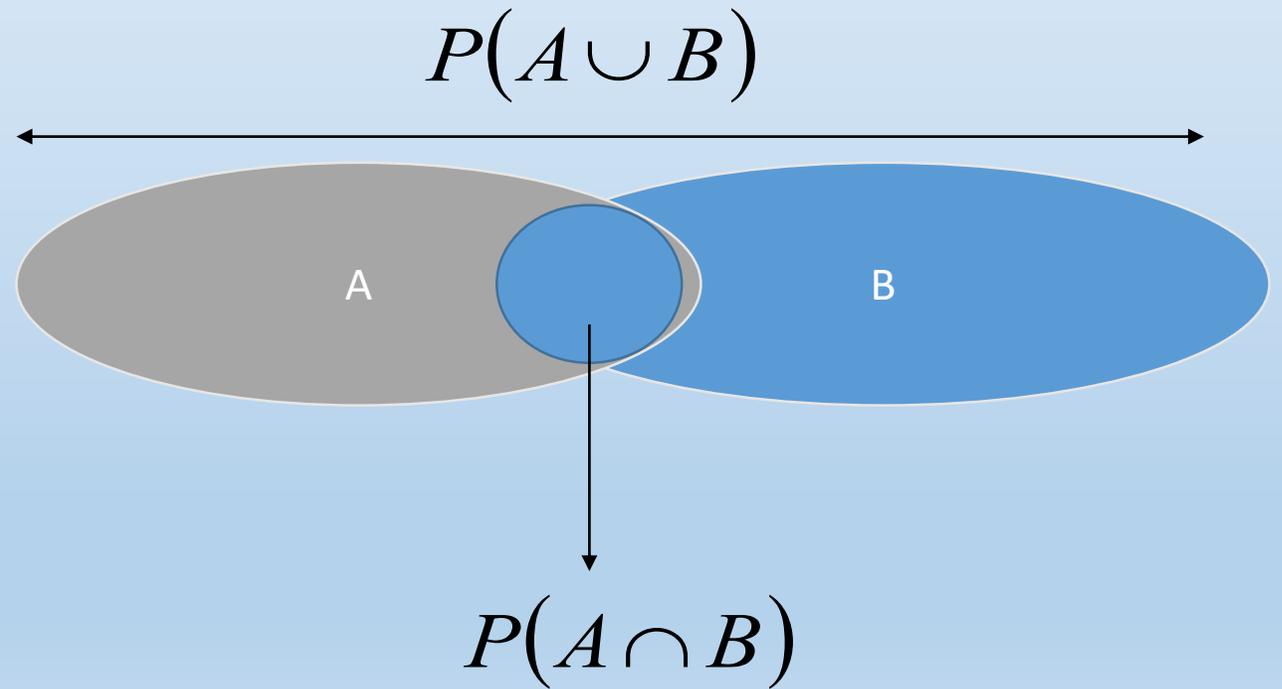
Per la probabilità condizionata, si fa riferimento **non all'insieme** di tutti i casi possibili, ma al **sottoinsieme** dei casi che verificano l'evento «X»: su tale sottoinsieme va calcolata la probabilità di estrarre una fiala difettosa.

Tale probabilità viene indicata con la notazione $P(D|X)$, che si legge “**probabilità condizionata a X**” e si può ottenere dal rapporto tra il numero di fiale che sono **sia difettose sia prodotte da X** ed il numero di fiale prodotte da X:

$$P(D | X) = \frac{P(D \cap X)}{P(X)}$$

$$P(A \cap B) = P(A \text{ e } B)$$

$$P(A \cup B) = P(A \text{ o } B)$$



FIALE	MACCHINE			X o Y o Z
	X	Y	Z	
\bar{d}	920	2760	5820	9500
d	80	240	180	500
d o \bar{d}	1000	3000	6000	10000

$$P(X) = \frac{1000}{10000} = 0,1$$

$$P(D) = \frac{500}{10000} = 0,05$$

$$P(X|D) = \frac{80}{500} = 0,16$$

$$P(D|X) = \frac{P(D \cap X)}{P(X)} = \frac{80}{1000} = 0,08$$

indipendentemente da quale sarà tra i due l'evento condizionante si ottiene lo stesso risultato per l'intersezione:

$$P(D \cap X) = P(X) * P(D|X) = 0,1 * 0,08 = 0,008$$

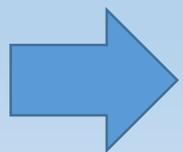
$$P(D \cap X) = P(D) * P(X|D) = 0,05 * 0,16 = 0,008$$

Dati due eventi A e B si ha che:

Bayes' formula:
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Ci permette di collegare le probabilità dei due distinti eventi: $A|B$ e $B|A$

Queste identità conseguono direttamente dalle definizioni di probabilità e probabilità condizionata.



Utilizzato in epidemiologia per calcolare la probabilità che un individuo abbia una malattia, dato che risulta **positivo** in un test di **screening**. Alla base della **impostazione bayesiana** dell'inferenza statistica.

Probabilità condizionata per eventi indipendenti

Quando A e B sono eventi **indipendenti** si ha che $P(B|A) = P(B)$

$$P(B|A) = \frac{P(A \text{ e } B)}{P(A)} = \frac{P(A)P(B)}{P(A)}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A)P(B)}{P(A)}$$

La conoscenza di A **non influisce** sulla probabilità che B si verifichi quando A e B sono eventi indipendenti.

Consideriamo due pazienti completamente diversi, che non si conoscono e che vivono in luoghi molto distanti tra loro.

Evento A: Il primo paziente si rompe un braccio giocando a basket.

Evento B: Il secondo paziente si ammala di influenza.

Sapere che il primo paziente si è rotto un braccio non ci fornisce **alcuna informazione aggiuntiva** sulla probabilità che il secondo paziente si ammali di influenza. Questi due eventi sono completamente scollegati e si verificano indipendentemente l'uno dall'altro.

A = Sindrome di Down

B = test positivo (alla futura mamma)

- $P(A) = 1/270$: 1 su 270 bambini (da donne ≥ 35 anni) hanno la sindrome di Down (0.4%)
- $P(B|A) \cong 0.89$: circa l'89% dei bambini con sindrome di Down erano positivi al test fatto in gravidanza
- $P(\text{not } B|\text{not } A) \cong 0.75$: circa il 75% dei bambini che non hanno la sindrome di Down erano negativi al test fatto in gravidanza
- $P(B|\text{not } A) \cong 0.25 \rightarrow (1 - P(\text{not } B|\text{not } A))$

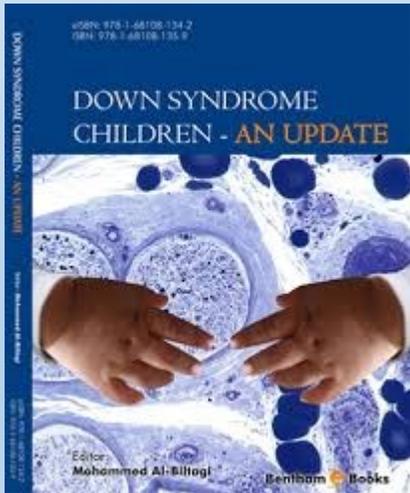
Se una donna 35+ effettua il test e risulta positivo, quale è la probabilità che il bimbo abbia la sindrome di Down ?

A = Sindrome di Down

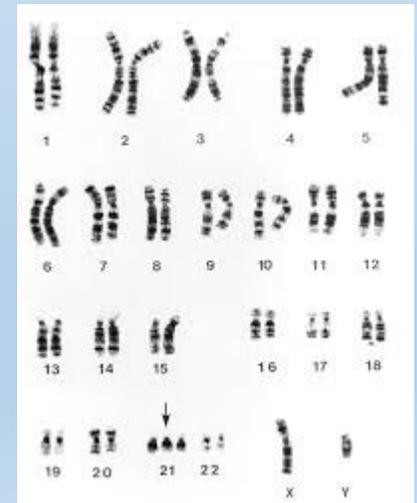
B = test positivo (alla futura mamma)

$$P(B) = P(B \& A) + P(B \& \text{not } A) = P(B|A)P(A) + P(B|\text{not } A)P(\text{not } A) \\ \cong .89*(1/270) + .25*(269/270) = .252$$

$$P(A|B) = P(B|A)P(A)/P(B) \cong .89(1/270)/.252 = .013=1.3\%$$



Invece di una possibilità inferiore a 4 su 1000, la probabilità di Down è maggiore di 1 su 100 se il test è positivo.



Test diagnostici

Per giungere ad una diagnosi il medico ipotizza un insieme di alternative. Tenta poi di ridurle escludendo progressivamente malattie specifiche.

In altri casi, il medico ha una forte convinzione che il malato sia affetto da una specifica malattia e cerca la conferma della sua ipotesi diagnostica.

Data una particolare diagnosi, un buon **test** dovrebbe indicare se la malattia è *improbabile o probabile*.

In pratica, è importante ricordare che un test diagnostico è utile solo se **il risultato influenza in modo rilevante** il trattamento del paziente.

Supponiamo di voler valutare la bontà di un test diagnostico che fornisce una **risposta dicotomica**: positiva o negativa, rispetto alla presenza o meno di una certa patologia.

Ci chiediamo:

(a) Se la malattia è presente qual è la probabilità che il test sia positivo?
-> questa domanda introduce il concetto di **sensibilità** di un test.

(b) Se la malattia è assente, qual è la probabilità di un risultato negativo?
-> questa domanda introduce il concetto di **specificità** di un test.

Ovviamente si può rispondere con precisione a queste domande solo se si conosce la diagnosi “vera”.

Ad esempio, la risposta *certa* può venire da una biopsia oppure da una procedura rischiosa e costosa come l'angiografia nel caso delle cardiopatie. O in altri casi si può ricorrere al parere di un "esperto". Questo tipo di risposte rappresentano ciò che comunemente è definito il **"gold standard"**.

Sospetta cardiomiopatia ischemica		Malattia coronarica (coronarografia)		Totale
		Presente (D+)	Assente (D-)	
Prova da sforzo Test	Positiva (T+)	815 (a)	115 (b)	930
	Negativa (T-)	208 (c)	327 (d)	535
Totale		1023	442	1465

La *prevalenza* di malattia coronarica in questi pazienti è $1023/1465=0.70$ o 70%.

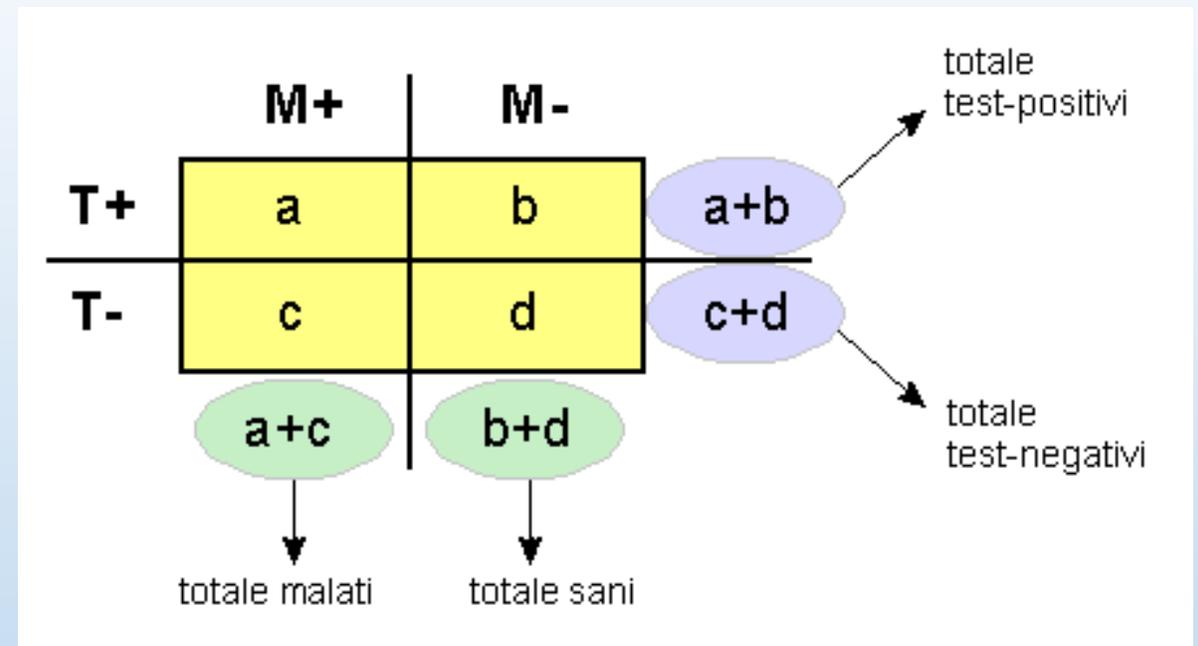
Possiamo scrivere come $P(\mathbf{D+})=0.70$

	Paziente Malato (M^+)	Paziente Sano (M^-)	Totale
Test Positivo (T^+)	VP (Veri Positivi)	FP (Falsi Positivi)	VP + FP
Test Negativo (T^-)	FN (Falsi Negativi)	VN (Veri Negativi)	FN + VN
Totale	VP + FN	FP + VN	

$$\text{accuratezza} = \frac{VP + VN}{VP + FN + FP + VN}$$

Proporzione di diagnosi corrette

		Malattia coronarica		Totale
		Presente (D+)	Assente (D-)	
Prova da sforzo	Positiva (T+)	815 (a)	115 (b)	930
	Negativa (T-)	208 (c)	327 (d)	535
Totale		1023	442	1465



La sensibilità risponde alla domanda: « quanti dei pazienti *malati* sottoposti al test, sono risultati positivi? »

La *sensibilità* di un test è quindi la proporzione di coloro che sono affetti dalla malattia e sono positivi al test.
 $a/(a+c)=815/1023=0.80$ o 80%.
sensibilità= $P(T+ | D+)$

La specificità risponde alla domanda: « quanti dei pazienti *sani* sottoposti al test, sono risultati negativi? »

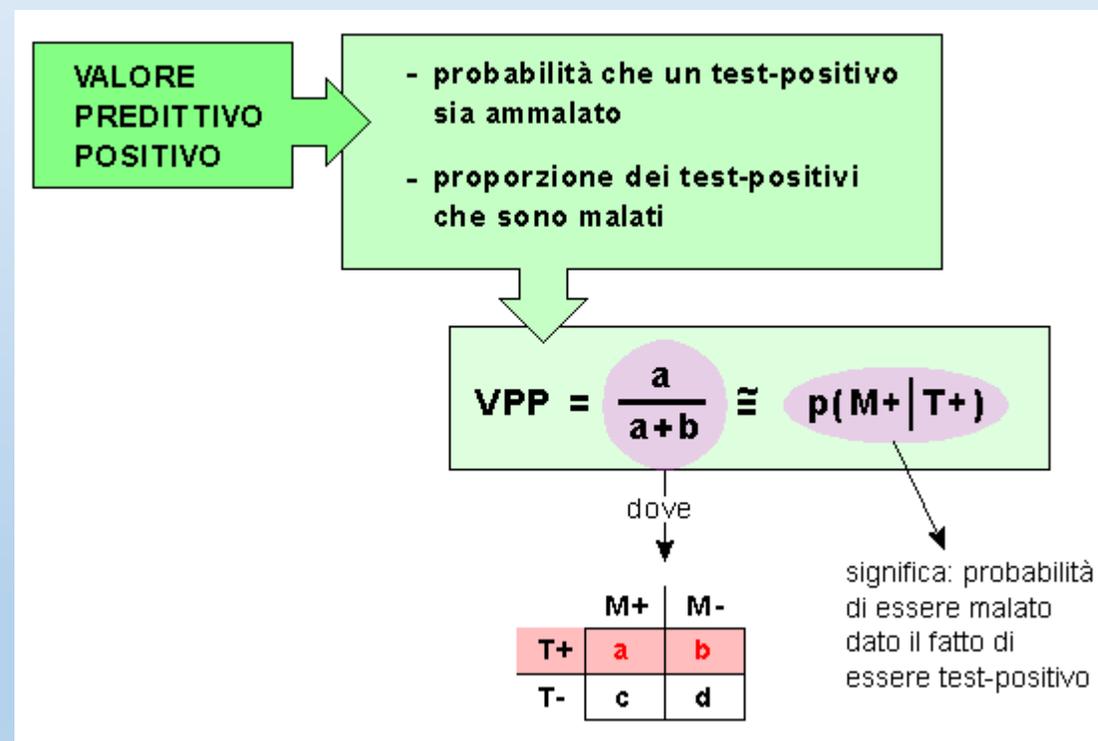
La *specificità* del test è la proporzione di coloro che non sono affetti dalla malattia e sono negativi al test:
 $d/(b+d)=327/442=0.74$ o 74%.
specificità= $P(T- | D-)$

Un medico vede un paziente affetto da dolore toracico compatibile con angina. Conoscendo la prevalenza della malattia, il medico pensa che il paziente sia affetto da malattia coronarica con una probabilità del 70%.

Il paziente viene poi sottoposto ad una prova da sforzo, ed il risultato è positivo.

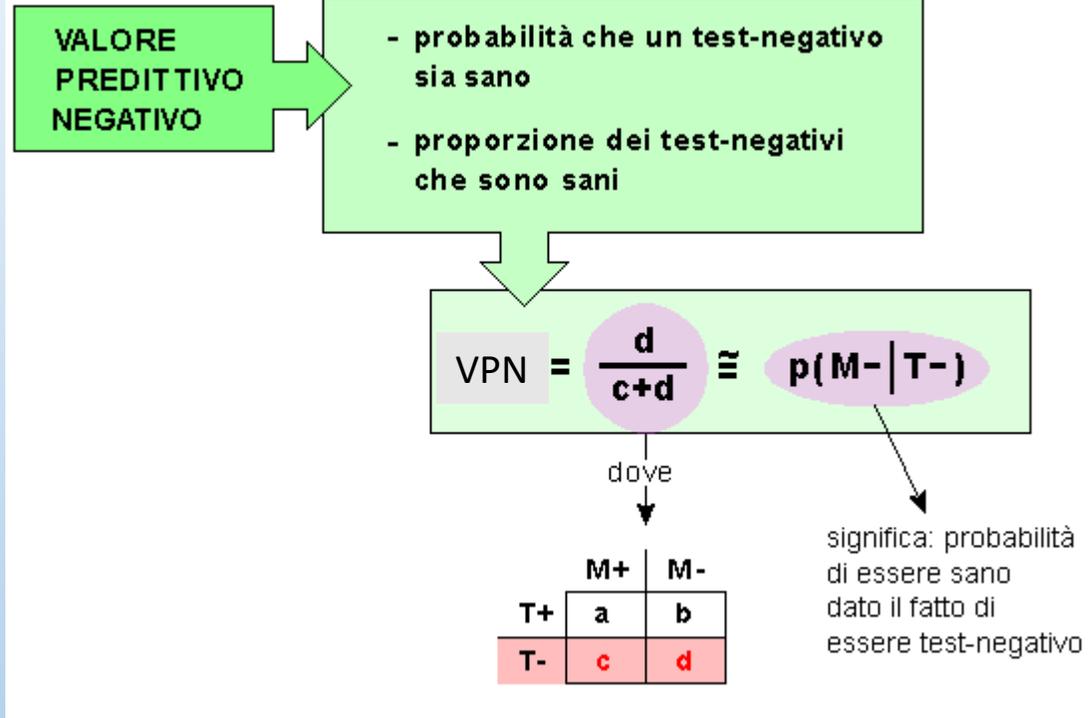
Di quanto il risultato positivo modifica le probabilità? (questo è quello che serve!!!)

		Malattia coronarica		Totale
		Presente (D+)	Assente (D-)	
Prova da sforzo	Positiva (T+)	815 (a)	115 (b)	930 (a+b)
	Negativa (T-)	208 (c)	327 (d)	535 (c+d)
Totale		1023	442	1465



Il *valore predittivo* di un test *positivo* è $P(D+ | T+)$: $VPP = 815/930 = 0.88$

		Malattia coronarica		Totale
		Presente (D+)	Assente (D-)	
Prova da sforzo	Positiva (T+)	815 (a)	115 (b)	930 (a+b)
	Negativa (T-)	208 (c)	327 (d)	535 (c+d)
Totale		1023	442	1465



Il *valore predittivo* di un test *negativo* è $P(D- | T-)$: $VPN=327/535=0.61$

In base quindi all'esito del test, la precedente stima di 0.70 deve essere *corretta verso l'alto* tenendo conto della probabilità di malattia "dato che" il test è positivo -> il valore predittivo del test positivo $815/930=0.88$.

Qual è la relazione tra VPP $P(D+ | T+)$ e la sensibilità $P(T+ | D+)$?

Il teorema di Bayes !!!

Dati due eventi A e B, la probabilità **congiunta** di A e B è data dalla regola della moltiplicazione:

$$P(A \text{ e } B) = P(A) * P(B | A)$$

A= test da sforzo positivo=T+

B= coronarografia positiva=D+

$$P(\mathbf{T+ e D+})=P(\mathbf{T+ | D+}) * P(\mathbf{D+})$$

Sensibilità * prevalenza

A= test da sforzo positivo=T+
B= coronarografia positiva=D+

Poiché in generale $P(A \text{ e } B)=P(B \text{ e } A)$ otteniamo che:

$$P(A | B) * P(B) = P(B | A) * P(A)$$

Questo passaggio ci porta al teorema di Bayes:

$$P(B | A) = [P(A | B) * P(B)] / P(A)$$

E' come dire:

$$P(\mathbf{D+ | T+}) = \text{valore predittivo del test positivo} =$$

$$[\text{sensibilità} * \text{prevalenza}] / \text{probabilità di un test positivo} =$$

$$[P(\mathbf{T+ | D+}) * P(\mathbf{D+})] / P(\mathbf{T+})$$

$$\begin{aligned} P(D+ | T+) &= \text{valore predittivo del test} \\ &\text{positivo=} \\ &[\text{sensibilità} * \text{prevalenza}] / \text{probabilità} \\ &\text{di un test positivo=} \\ &[P(T+ | D+) * P(D+)] / P(T+) \end{aligned}$$

Se abbiamo dei dati di letteratura sulla sensibilità e specificità di un test diagnostico, e siamo in grado di avere una stima della prevalenza possiamo ricavarci il valore predittivo positivo come:

$$PPV = \frac{\textit{sensitivity} \times \textit{prevalence}}{\textit{sensitivity} \times \textit{prevalence} + (1 - \textit{specificity}) \times (1 - \textit{prevalence})}$$

Qual è la relazione tra VPN $P(D^- | T^-)$, la specificità $P(T^- | D^-)$ e la prevalenza $P(D^+)$?

specificità

1- prevalenza

$$VPN = \frac{P(T^- | D^-) * (1 - P(D^+))}{P(T^-)}$$

Equivalente a :

$$NPV = \frac{\textit{specificity} \times (1 - \textit{prevalence})}{(1 - \textit{sensitivity}) \times \textit{prevalence} + \textit{specificity} \times (1 - \textit{prevalence})}$$

Notiamo che all'aumentare della prevalenza il valore predittivo negativo diminuisce (in una popolazione dove una certa malattia è molto diffusa è meno probabile che un test negativo indichi che il paziente sia realmente sano).

$P(D+ | T+)$ =valore predittivo del test
positivo=0.88

$P(T+ | D+)$ =sensibilità= 0.80

$P(D+)$ =prevalenza=0.70

$P(T+)$ =prob. test positivo=930/1465=0.63

Teorema di Bayes:

$P(D+ | T+) = [P(T+ | D+) * P(D+)] / P(T+)$

$(0.80 * 0.70) / 0.63 = 0.56 / 0.63 = \mathbf{0.88}$

$P(D+ | T+)$ =probabilità a posteriori

$P(D+)$ =probabilità a priori

		Malattia coronarica		Totale
		Presente (D+)	Assente (D-)	
Prova da sforzo	Positiva (T+)	815 (a)	115 (b)	930 (a+b)
	Negativa (T-)	208 (c)	327 (d)	535 (c+d)
Totale		1023	442	1465

Una volta eseguito il test, Sensibilità e Specificità perdono importanza. Diventano importanti le probabilità *post-test*. E' auspicabile avere un'idea della quota di soggetti realmente ammalati (**positivi veri**, cella a) sul totale dei soggetti risultati positivi al test (celle a+b).

Analogamente, è bene conoscere la quota di soggetti realmente sani, (**negativi veri**, cella d) sul totale dei negativi al test (celle c+d).

La **prevalenza** di una malattia è 1 su 1000 e disponiamo di un test che può diagnosticarla con una sensibilità del 100% ed una specificità del 95%.

Qual è la probabilità che una persona sia affetta da malattia, in presenza di un test positivo?

Per calcolare la probabilità di un test positivo, partiamo da una popolazione di 1000 persone, di cui 1 è affetta dalla malattia. Il test sicuramente individuerà quella persona (sensibilità del 100%), ma sarà positivo anche per il 5% delle 999 persone sane. Quindi, il numero totale di test positivi sarà: $1 + 0.05 * 999 = 50.95$.

$$P(T+) = 50.95 / 1000 = 0.05095$$

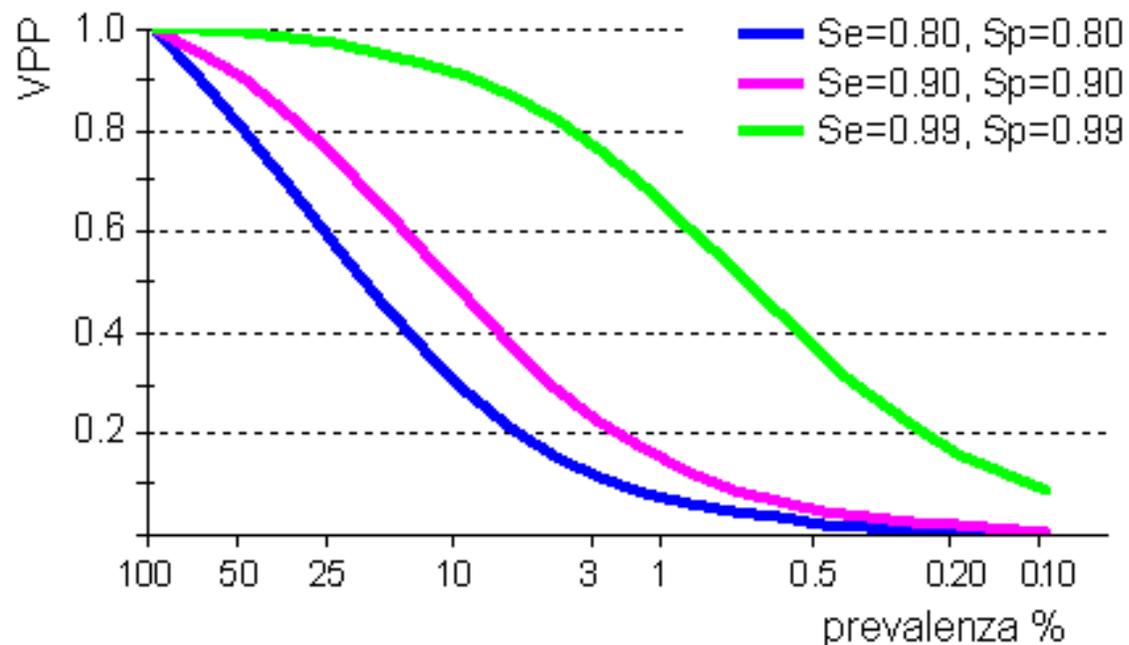
$$P(D+) = 1 / 1000 = 0.001$$

sensibilità=1

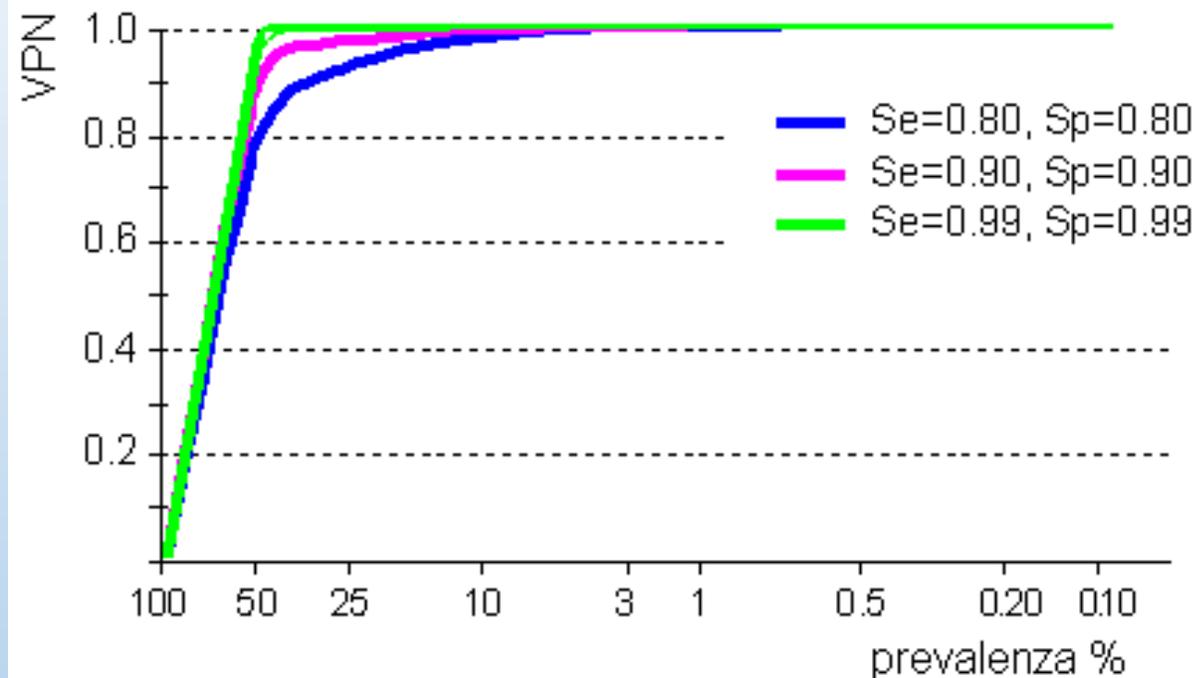


$$P(D+ | T+) = [P(T+ | D+)P(D+)] / P(T+) = [1 * 0.001] / 0.05095 = \mathbf{0.02}$$

Valore predittivo positivo (VPP) in rapporto alla sensibilità e specificità del test ed alla prevalenza della malattia



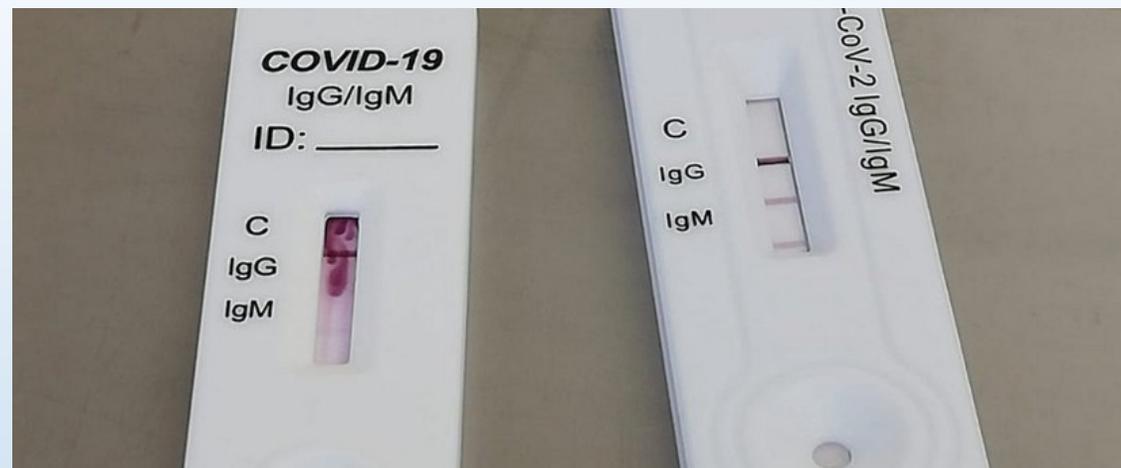
Valore predittivo negativo (VPN) in rapporto alla sensibilità e specificità del test, ed alla prevalenza della malattia



L'utilità di un test dipende dalla prevalenza della malattia. Un test è utile se viene modificata in modo rilevante la probabilità pre-test. Se una malattia è molto rara oppure molto frequente, il test ha una utilità discutibile.



Sono guarito?



Come si effettuava la verifica di guarigione dal Covid-19?

Il paziente doveva sottoporsi a **2 tamponi** a distanza di 24 ore l'uno dall'altro.

Se entrambi risultavano negativi, si poteva affermare con una "certa sicurezza" la scomparsa del virus.

Perché due ?

Tabella dati su un test sierologico

	Malato (M)	Sano (S)	Totale
Test Positivo (T^+)	17	2	19
Test Negativo (T^-)	3	48	51
Totale	20	50	70

	Malato (M)	Sano (S)	Totale
Test Positivo (T^+)	17/70	2/70	19/70
Test Negativo (T^-)	3/70	48/70	51/70
Totale	20/70	50/70	1

	Malato (M)	Sano (S)	
Test positivo (T^+)	$P(M \text{ e } T^+)$	$P(S \text{ e } T^+)$	$P(T^+)$
Test negativo (T^-)	$P(M \text{ e } T^-)$	$P(S \text{ e } T^-)$	$P(T^-)$
	$P(M)$	$P(S)$	

- Il test diagnostico utilizzato per la diagnosi del Covid-19 è il test molecolare con metodo Real Time PCR per SARS-CoV-2 indicato dall'OMS (il cosiddetto tampone)
- Il test ha (valori arrotondati) di sensibilità del **95%** e di specificità del **95%**

Formalizziamo le caratteristiche del test:

M = il paziente è malato

S = il paziente è sano

T+ = il test è positivo

T- = il test è negativo

$$\text{Sensibilità} = \frac{P(M \text{ e } T^+)}{P(M)} = P(T^+ | M) = 0.95$$
$$\text{Specificità} = \frac{P(S \text{ e } T^-)}{P(S)} = P(T^- | S) = 0.95$$

A noi interessa:

$P(S|T-)$ =probabilità di **essere sano** essendo risultato **negativo** al test

Sappiamo che :

$$P(T+|M) = P(T-|S) = 0.95$$

$$P(T-|M) = P(T+|S) = 0.05$$

Dobbiamo assegnare un valore ipotizzato alla prevalenza del COVID: $P(M)$

- Assegniamo: $P(M) = 0.2$ (20%)
- Quindi $P(S) = 1 - P(M) = 0.8$ (80%)
- Noi vogliamo **$P(S|T-)$**

	Malato	Sano	
Test positivo	$P(M e T^+)$	$P(S e T^+)$	$P(T^+)$
Test negativo	$P(M e T^-)$	$P(S e T^-)$	$P(T^-)$
	$P(M)$	$P(S)$	

$$P(S|T^-) = \frac{P(S e T^-)}{P(T^-)}$$

$$\begin{aligned}
 P(S|T^-) &= \frac{P(S e T^-)}{P(T^-)} = \frac{P(T^-|S) * P(S)}{P(S e T^-) + P(M e T^-)} = \\
 &= \frac{P(T^-|S) * P(S)}{P(T^-|S) * P(S) + P(T^-|M) * P(M)} = \frac{0.95 * 0.8}{0.95 * 0.8 + 0.05 * 0.2} = 0.987
 \end{aligned}$$

con una probabilità (a posteriori) del **98.7%** sono guarito se il mio tampone è negativo

E perché allora effettuare **due** tamponi?

Ho una probabilità del 98.7% di essere guarito se il mio primo tampone è negativo...

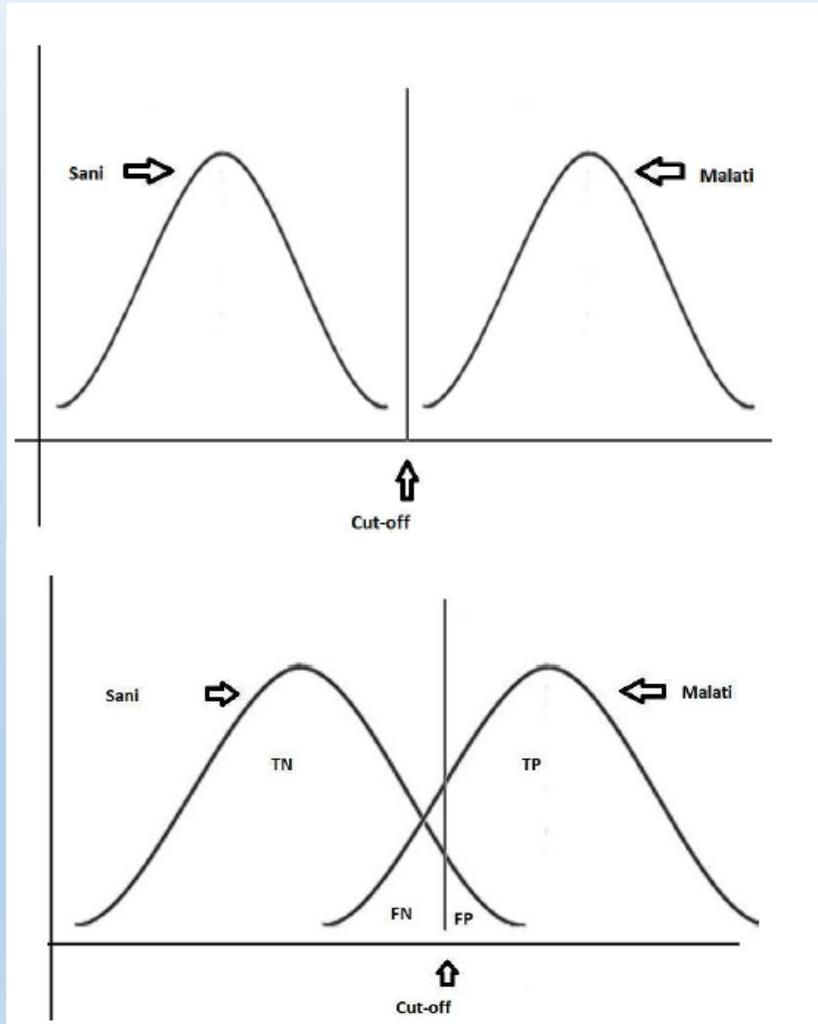
Rifaccio un altro tampone e calcoliamo: $P(S|T_1^- \text{ e } T_2^-)$

Si usa sempre il **teorema di Bayes** ma con 0.987 come **probabilità iniziale**: ossia $P(S|T_1^-)$ al posto di $P(S)$:

$$\begin{aligned}
 P(S|T_1^-, T_2^-) &= \frac{P(T_2^-|S) * P(S|T_1^-)}{P(T_2^-|S) * P(S|T_1^-) + P(T_2^-|M) * P(M|T_1^-)} = \\
 &= \frac{0.95 * 0.987}{0.95 * 0.987 + 0.05 * 0.013} = \mathbf{0.999}
 \end{aligned}$$



I test diagnostici ideali (i cosiddetti **gold standard**) discriminano perfettamente i malati dai sani: gli individui vengono classificati con assoluta certezza come affetti o non affetti dalla malattia di interesse.



Quando un determinato *test* diagnostico non discrimina in maniera netta i malati dai sani, è necessario calcolare **il grado di incertezza** della classificazione.

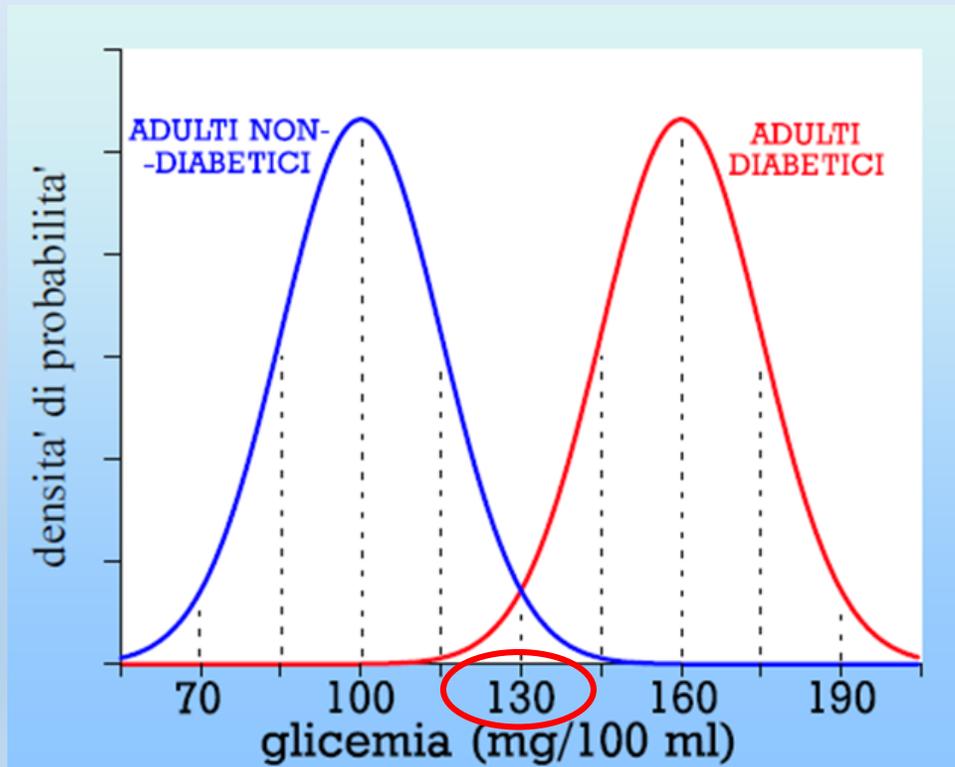
Se il risultato del test diagnostico è una **variabile binaria** (affetto/non affetto), è sufficiente calcolare:

- la sensibilità
- la specificità
- i valori predittivi (positivo e negativo)

Se il risultato del test è invece una **variabile continua numerica**, occorre utilizzare l'analisi della **curva ROC** (acronimo di *Receiver Operating Characteristics*).

La curva ROC è una tecnica statistica che misura l'accuratezza di un test diagnostico lungo tutto il **range** dei valori possibili.

La curva ROC rappresenta il metodo d'elezione per validare un test diagnostico.



La curva **ROC** permette anche di identificare il valore soglia **ottimale** (il cosiddetto *best cut-off*), cioè il valore del test che massimizza la differenza tra i veri positivi e i falsi positivi...

Per ogni possibile valore del biomarcatore si definisce un cut-off (c_i) e si calcola la tabella di contingenza (supponiamo qui che valori crescenti indicano aumento del rischio):

	MALATI	SANI
Marker $\geq c_1$	VP	FP
Marker $< c_1$	FN	VN



Sensibilità(1)
Specificità(1)

	MALATI	SANI
Marker $\geq c_2$	VP	FP
Marker $< c_2$	FN	VN



Sensibilità(2)
Specificità(2)

	MALATI	SANI
Marker $\geq c_3$	VP	FP
Marker $< c_3$	FN	VN

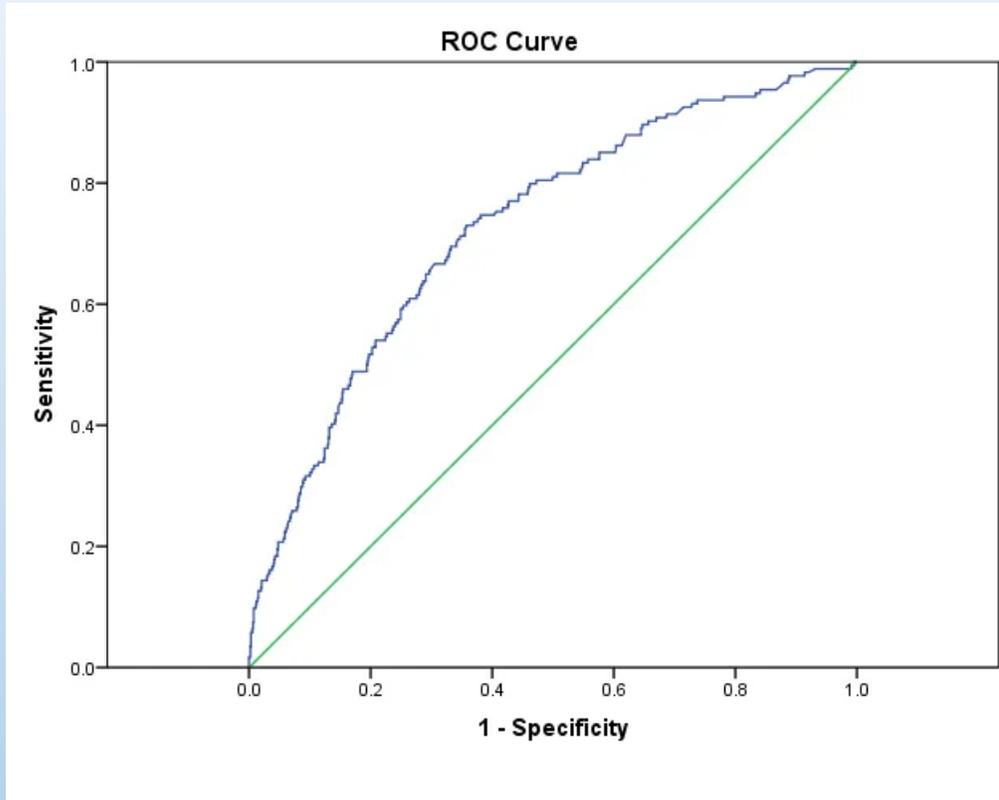


Sensibilità(3)
Specificità(3)

Si ripete questa operazione *su ogni possibile valore* del marcatore e si ottiene una *distribuzione* di valori di sensibilità e specificità

Che cos'è la poi Curva ROC?

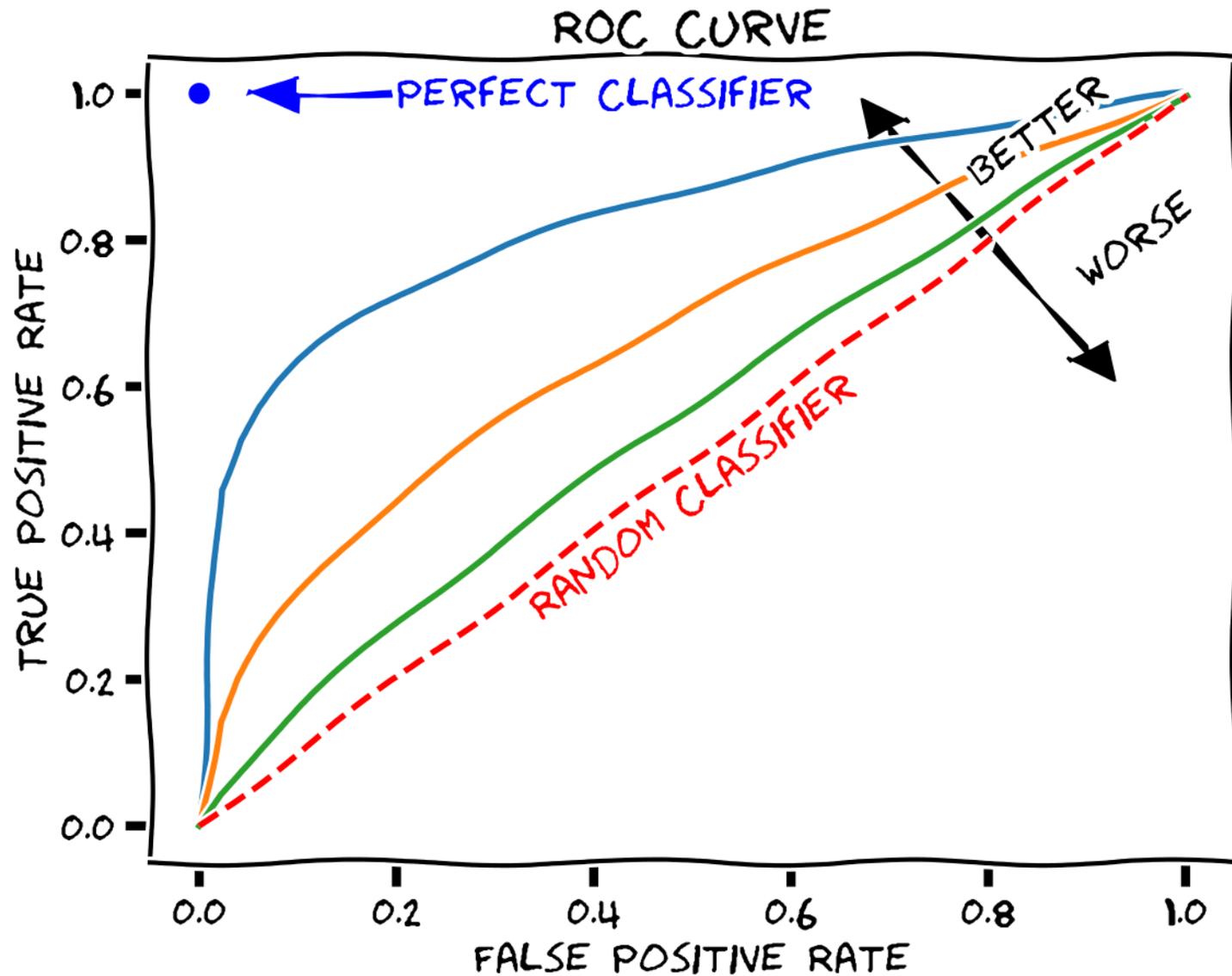
è un grafico in cui vengono riportati (per diversi valori del test diagnostico) in ordinata la **Sensibilità** e in ascissa **1-Specificità**:



Qual è l'utilità della curva ROC?

- permette di scegliere il **migliore*** cut-off in un test diagnostico
- permette di scegliere il **migliore** tra due test diagnostici: quello con l' *Area Under the Curve* (AUC) maggiore

*Il cut-off sarebbe **ottimale** se massimizzasse **contemporaneamente** sia la sensibilità che la specificità. Questo tuttavia non è possibile: infatti al crescere della specificità, diminuisce il valore di falsi positivi, ma si aumentano i falsi negativi, il che comporta una diminuzione della sensibilità. Si può quindi osservare che esiste un **trade-off** tra i due indici.



Nel caso in cui il nuovo *test* non discrimina i malati dai sani, la curva ROC ha un'area (AUC) di 0.5 (o 50%) che coincide con l'area sottostante la diagonale del grafico (Reference Line).

Si considera adeguato un *test* diagnostico con un'area sotto la curva $\geq 80\%$.

L'area sotto la curva può assumere valori compresi tra 0.5 e 1.0. Tanto maggiore è l'area sotto la curva tanto maggiore è il potere discriminante del *test*.

Per l'**interpretazione** dei valori dell'area sottostante la curva ROC è possibile riferirsi alla classificazione proposta da Swets (1988)*:

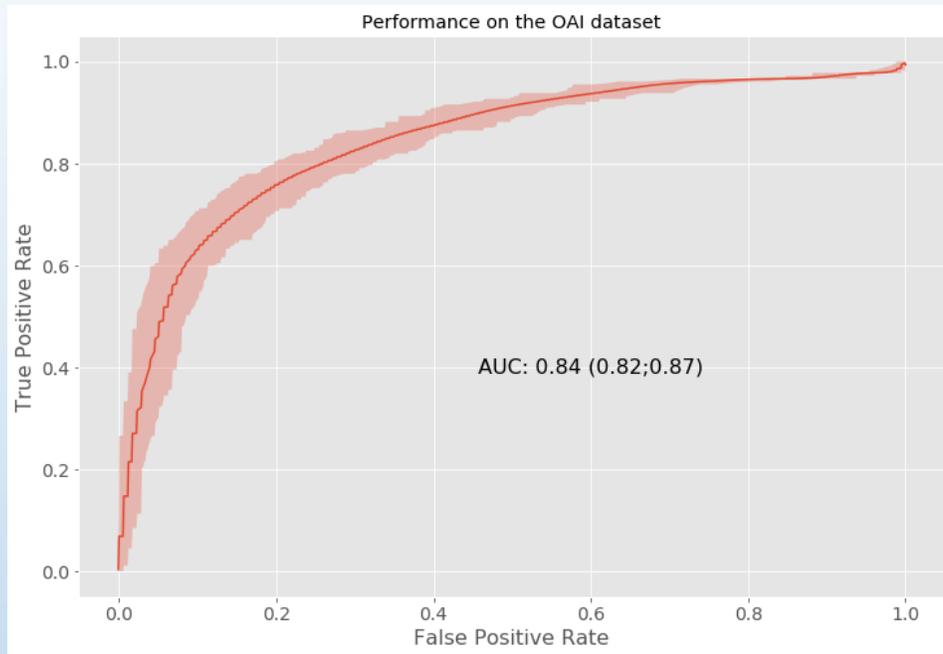
$AUC=0.5$ il *test* non è informativo

$0.5 < AUC \leq 0.7$ il *test* è poco accurato

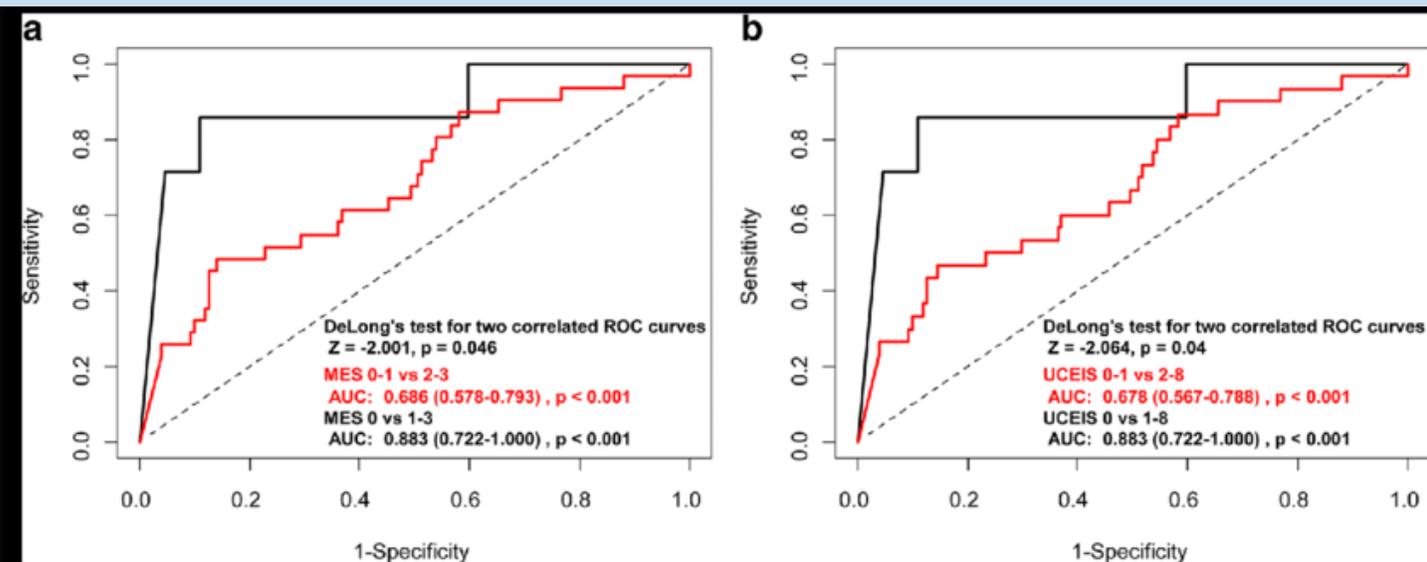
$0.7 < AUC \leq 0.9$ il *test* è moderatamente accurato

$0.9 < AUC < 1.0$ il *test* è altamente accurato

$AUC=1$ *test* perfetto



Non entriamo adesso nel dettaglio del calcolo degli **intervalli di confidenza** ma è un dato che va sempre riportato nei lavori scientifici.



Infine, si possono eseguire dei **test statistici** di ipotesi per confrontare due test diagnostici: ad esempio il De Long test:

$$H_0: AUC_1 = AUC_2$$

$$H_1: AUC_1 \neq AUC_2$$