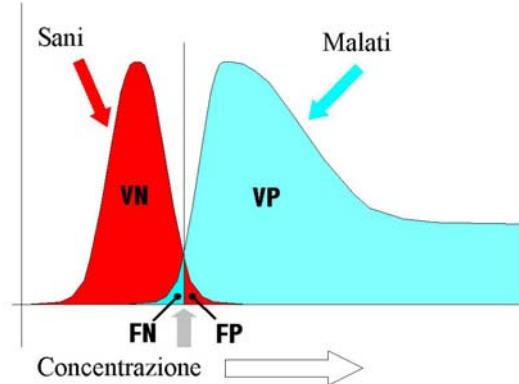
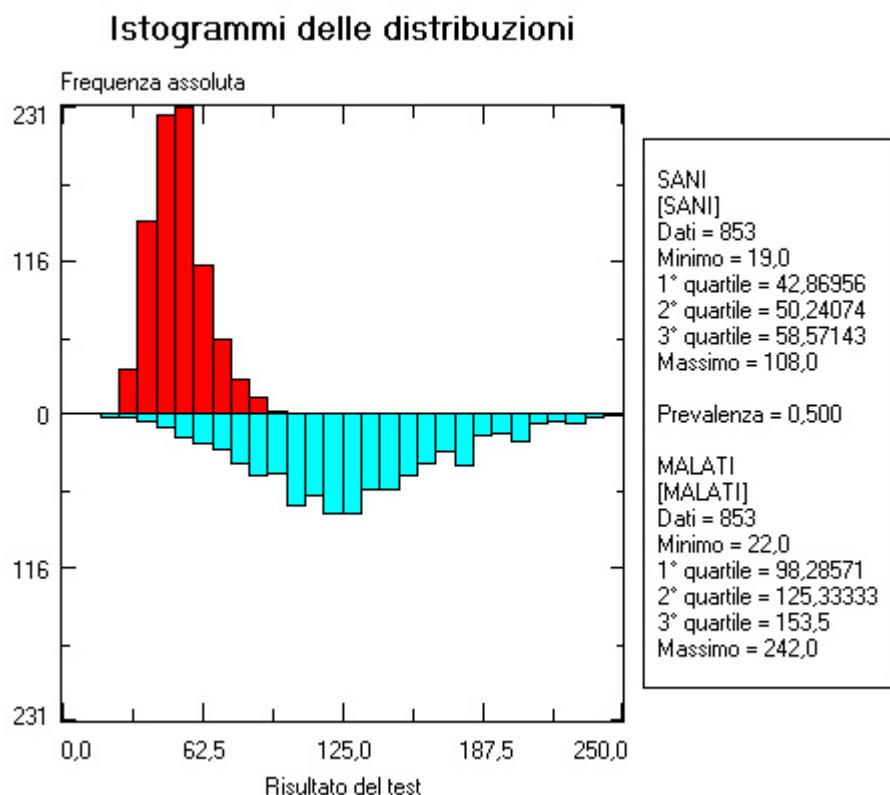


7a. Un esempio di analisi bayesiana dei dati mediante Ministat

Consideriamo una situazione nella quale la concentrazione dell'analita aumenti nella malattia, e nella quale la soglia tra sani e malati è fissata in un punto intermedio e tale per cui si abbiano il massimo possibile di veri negativi (VN = soggetti sani classificati come tali dal test) e il massimo possibile di veri positivi (VP = soggetti malati classificati come tali dal test), ma anche una quota di falsi positivi (FP = soggetti sani classificati come malati dal test) e una quota di falsi negativi (FN = soggetti malati classificati come sani dal test).



Per una completa analisi bayesiana dei dati, che elimina il fastidio di dovere effettuare i calcoli semplici ma lunghi che essa comporta, è possibile utilizzare software statistici biomedici, come Ministat [1], cui si riferiscono le immagini del seguente esempio simulato. Si suppongo di avere a disposizione i dati della determinazione della concentrazione nel siero di un dato analita, relativi a 853 soggetti sani e a 853 soggetti affetti da una malattia che provoca un patologia in aumento, un patologia cioè caratterizzata da un aumento della concentrazione dell'analita (nel caso di una patologia in diminuzione, cioè caratterizzata da una diminuzione della concentrazione dell'analita, bisogna semplicemente immaginare una situazione speculare a quella descritta). Le distribuzioni dei valori nei soggetti sani e nei soggetti malati sono poste a confronto nella figura seguente.



Nella tabella che segue è riportata l'analisi della casistica (VN = veri negativi, FP = falsi positivi, VP = veri positivi, FN = falsi negativi) per ciascuna delle classi in cui i dati sono stati suddivisi

Classe	da...	a...	Sani		Malati	
			VN	FP	VP	FN
1	0,0	8,33333	0	853	853	0
2	8,33333	16,66667	0	853	853	0
3	16,66667	25,0	2	851	850	3
4	25,0	33,33333	37	816	847	6
5	33,33333	41,66667	183	670	841	12
6	41,66667	50,0	408	445	831	22
7	50,0	58,33333	639	214	813	40
8	58,33333	66,66667	753	100	790	63
9	66,66667	75,0	809	44	763	90
10	75,0	83,33333	836	17	725	128
11	83,33333	91,66667	849	4	678	175
12	91,66667	100,0	852	1	632	221
13	100,0	108,33333	853	0	563	290
14	108,33333	116,66667	853	0	501	352
15	116,66667	125,0	853	0	425	428
16	125,0	133,33333	853	0	349	504
17	133,33333	141,66667	853	0	291	562
18	141,66667	150,0	853	0	233	620
19	150,0	158,33333	853	0	186	667
20	158,33333	166,66667	853	0	148	705
21	166,66667	175,0	853	0	120	733
22	175,0	183,33333	853	0	81	772
23	183,33333	191,66667	853	0	64	789
24	191,66667	200,0	853	0	48	805
25	200,0	208,33333	853	0	27	826
26	208,33333	216,66667	853	0	19	834
27	216,66667	225,0	853	0	13	840
28	225,0	233,33333	853	0	6	847
29	233,33333	241,66667	853	0	2	851
30	241,66667	250,0	853	0	0	853

Nella tabella che segue, viene riportata l'efficienza diagnostica del test (ultima colonna) in funzione dei livelli di soglia che è possibile fissare per discriminare tra sani e malati (prima colonna). Come si vede alla decima riga, la massima efficienza (0,84291) viene raggiunta se si considerano malati i soggetti con concentrazione uguale o superiore a 75, e sani i soggetti con concentrazione inferiore a 75. Si fa notare come sensibilità e specificità variano in modo opposto: fissando il valore soglia laddove si ha il massimo di sensibilità (1,0), il test risulterebbe totalmente aspecifico (darebbe cioè un risultato positivo anche a tutti i soggetti sani), mentre all'aumentare della specificità del test la sua sensibilità (la capacità cioè di identificare i malati) diminuisce.

Positivo se...	Sensibilità	Specificità	Efficienza
> 0,0	1,00000	0,00000	0,00000
> 8,33333	1,00000	0,00000	0,00000
> 16,66667	1,00000	0,00000	0,00000
> 25,0	0,99648	0,00234	0,00000
> 33,33333	0,99297	0,04338	0,03634
> 41,66667	0,98593	0,21454	0,20047
> 50,0	0,97421	0,47831	0,45252
> 58,33333	0,95311	0,74912	0,70223
> 66,66667	0,92614	0,88277	0,80891
> 75,0	0,89449	0,94842	0,84291
> 83,33333	0,84994	0,98007	0,83001
> 91,66667	0,79484	0,99531	0,79015
> 100,0	0,74091	0,99883	0,73974
> 108,33333	0,66002	1,00000	0,66002
> 116,66667	0,58734	1,00000	0,58734
> 125,0	0,49824	1,00000	0,49824
> 133,33333	0,40914	1,00000	0,40914
> 141,66667	0,34115	1,00000	0,34115
> 150,0	0,27315	1,00000	0,27315
> 158,33333	0,21805	1,00000	0,21805
> 166,66667	0,17351	1,00000	0,17351
> 175,0	0,14068	1,00000	0,14068
> 183,33333	0,09496	1,00000	0,09496
> 191,66667	0,07503	1,00000	0,07503
> 200,0	0,05627	1,00000	0,05627
> 208,33333	0,03165	1,00000	0,03165

> 216,66667	0,02227	1,00000	0,02227
> 225,0	0,01524	1,00000	0,01524
> 233,33333	0,00703	1,00000	0,00703
> 241,66667	0,00234	1,00000	0,00234
> 250,0	0,00000	1,00000	0,00000

Nella tabella che segue le caratteristiche del test il valore predittivo del test positivo e il valore predittivo del test negativo sono calcolati per ciascuno dei livelli soglia possibili..

Positivo se...	Sensibilità	Specificità	Valore predittivo del test	
			test positivo	test negativo
> 0,0	1,00000	0,00000	0,50000	###
> 8,33333	1,00000	0,00000	0,50000	###
> 16,66667	1,00000	0,00000	0,50000	###
> 25,0	0,99648	0,00234	0,49971	0,40000
> 33,33333	0,99297	0,04338	0,50932	0,86047
> 41,66667	0,98593	0,21454	0,55659	0,93846
> 50,0	0,97421	0,47831	0,65125	0,94884
> 58,33333	0,95311	0,74912	0,79163	0,94109
> 66,66667	0,92614	0,88277	0,88764	0,92279
> 75,0	0,89449	0,94842	0,94548	0,89989
> 83,33333	0,84994	0,98007	0,97709	0,86722
> 91,66667	0,79484	0,99531	0,99413	0,82910
> 100,0	0,74091	0,99883	0,99842	0,79404
> 108,33333	0,66002	1,00000	1,00000	0,74628
> 116,66667	0,58734	1,00000	1,00000	0,70788
> 125,0	0,49824	1,00000	1,00000	0,66589
> 133,33333	0,40914	1,00000	1,00000	0,62859
> 141,66667	0,34115	1,00000	1,00000	0,60283
> 150,0	0,27315	1,00000	1,00000	0,57909
> 158,33333	0,21805	1,00000	1,00000	0,56118
> 166,66667	0,17351	1,00000	1,00000	0,54750
> 175,0	0,14068	1,00000	1,00000	0,53783
> 183,33333	0,09496	1,00000	1,00000	0,52492
> 191,66667	0,07503	1,00000	1,00000	0,51949
> 200,0	0,05627	1,00000	1,00000	0,51448
> 208,33333	0,03165	1,00000	1,00000	0,50804
> 216,66667	0,02227	1,00000	1,00000	0,50563
> 225,0	0,01524	1,00000	1,00000	0,50384
> 233,33333	0,00703	1,00000	1,00000	0,50176
> 241,66667	0,00234	1,00000	1,00000	0,50059
> 250,0	0,00000	1,00000	###	0,50000

Quindi dato che, sulla base del criterio di massima efficienza del test, è stato scelto come livello di discriminazione tra sani e malati il valore di 75,0, è possibile fornire una misura quantitativa, e oggettiva, delle caratteristiche del test, che sono:

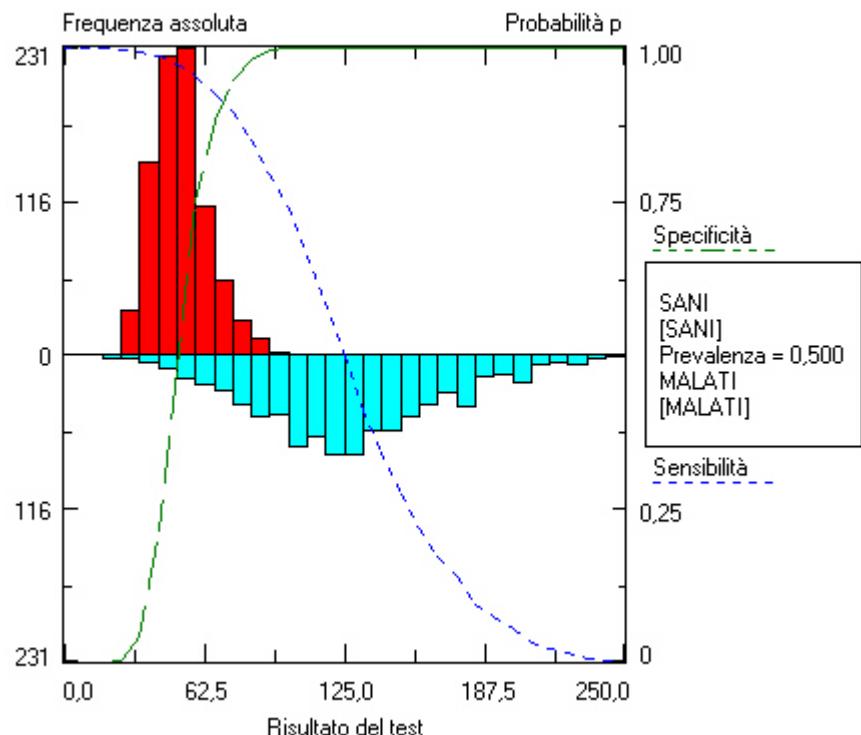
- ↳ sensibilità = 0,89449
- ↳ specificità = 0,94842
- ↳ valore predittivo del test positivo = 0,94548
- ↳ valore predittivo del test negativo = 0,89989

(chi preferisce i valori percentuali, deve solo moltiplicare tali valori per cento. I valori indicati con # non sono computabili per il numero di osservazioni effettuate).

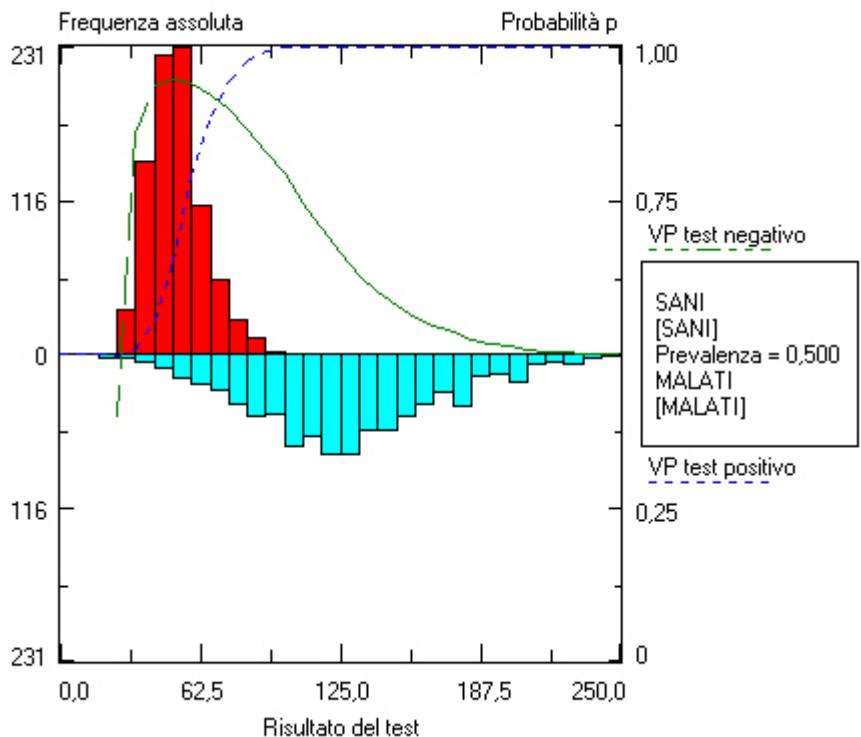
La rappresentazione grafica dell'andamento della sensibilità e della specificità in funzione del valore soglia prescelto è riportata nella prima delle due figure successive.

La rappresentazione grafica dell'andamento del valore predittivo del test positivo e del valore predittivo del test negativo in funzione del valore soglia prescelto è riportata nella seconda delle due figure successive.

Sensibilità e specificità del test

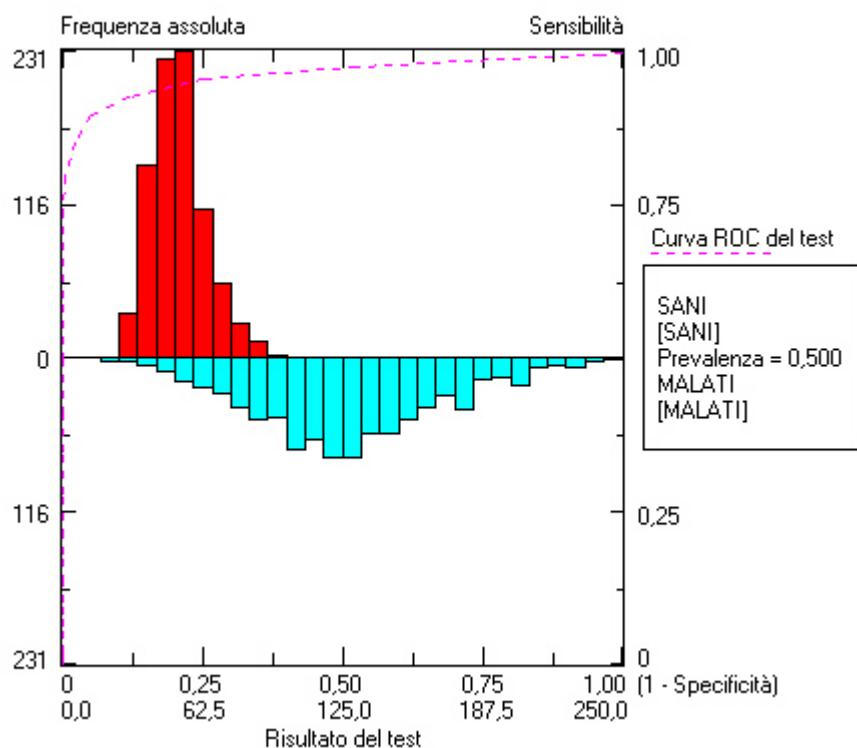


Valore predittivo del test

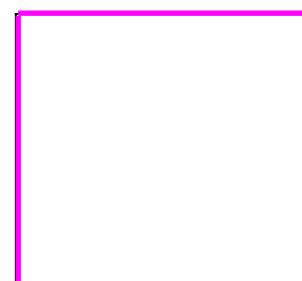
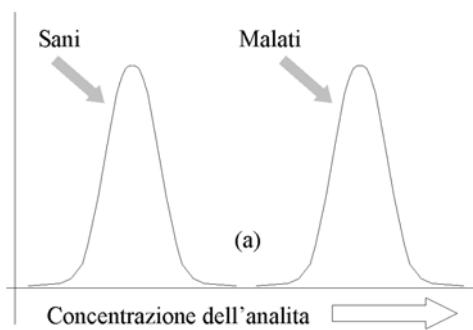


In fine riportando in un grafico in ordinate la sensibilità e in ascisse (1- specificità) si può ottenere la curva caratteristica (cirva ROC) del test

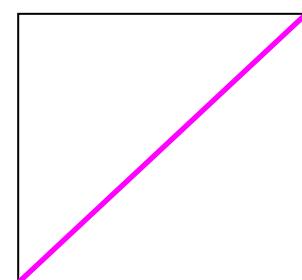
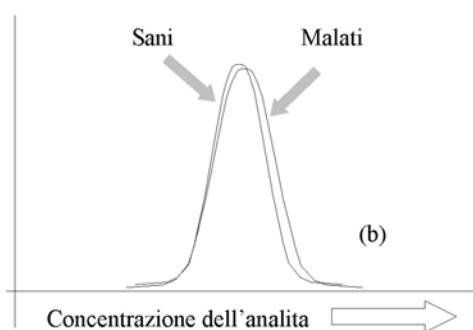
Curva ROC del test



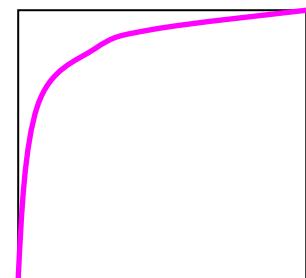
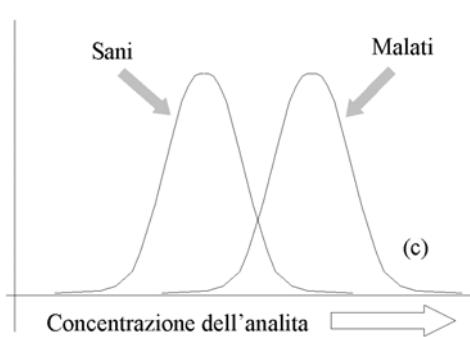
In pratica le curve ROC [1,2] riassumono in un unico indice le grandezze bayesiane illustrate analiticamente in precedenza, e forniscono una descrizione cumulativa del grado di sovrapposizione tra la distribuzione dei sani e dei malati, come indicato dalla corrispondenza tra questo e le curve ROC qui riportata.



Curva ROC di un test ideale



Curva ROC di un test "inutile"



Curva ROC di un test reale

[1] Besozzi M. http://www.bayes.it/download/Ministat_21.zip.

[2] Altman DG, Bland JM. *Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots*. BMJ 1994;309;188.