

Analisi dei dati diagnostici

Test diagnostico

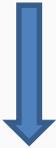


Affidabilità



Riproducibilità

o precisione



Concordezza

Validità

o accuratezza



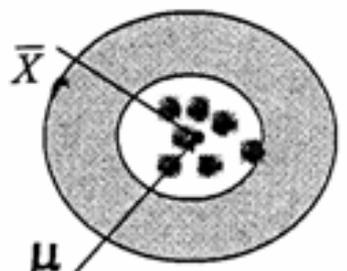
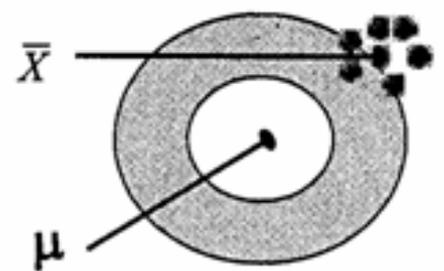
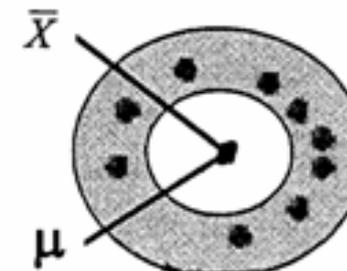
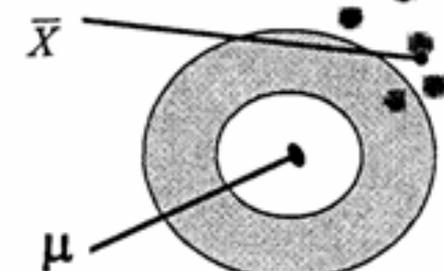
Sensibilità

Specificità

Valore predittivo positivo

Valore predittivo negativo

Precisione e accuratezza

| PRECISIONE ($x_i - \bar{x}$) | ACCURATEZZA ($\bar{x} - \mu$) | |
|-----------------------------------|--|---|
| | Elevata | Scarsa |
| Elevata |  |  |
| Scarsa |  |  |

Concordanza - Kappa di Cohen

Misura la concordanza specifica tra due osservatori (es. tra due patologi, tra due radiologi), tra due test diagnostici o intra-osservatore in tempi diversi.

Il vantaggio di questo indice è dato dalla correzione per la concordanza che può verificarsi solo per effetto del caso

Kappa di Cohen

Supponiamo che a due medici venga chiesto di esaminare 100 pap test

| | + | - | Med 2 |
|-------|----|----|-------|
| + | 52 | 8 | 60 |
| - | 8 | 32 | 40 |
| Med 1 | 60 | 40 | 100 |

Accordo osservato = $(52+32)/100 =$ **84%**

Possiamo dire che la concordanza è dell'84%?

Kappa di Cohen

Calcolo dell'accordo per caso

| | + | - | Med 2 |
|-------|------------------------|------------------------|-------|
| + | $(60 \times 60) / 100$ | $(60 \times 40) / 100$ | 60 |
| - | $(60 \times 40) / 100$ | $(40 \times 40) / 100$ | 40 |
| Med 1 | 60 | 40 | 100 |

| | + | - | Med 2 |
|-------|----|----|-------|
| + | 36 | 24 | 60 |
| - | 24 | 16 | 40 |
| Med 1 | 60 | 40 | 100 |

Accordo per caso = $(36 + 16) / 100 =$ **52%**

Kappa di Cohen

Calcolo di kappa

Accordo per caso= $(36+16)/100$ 52%

Accordo residuo non casuale 48%

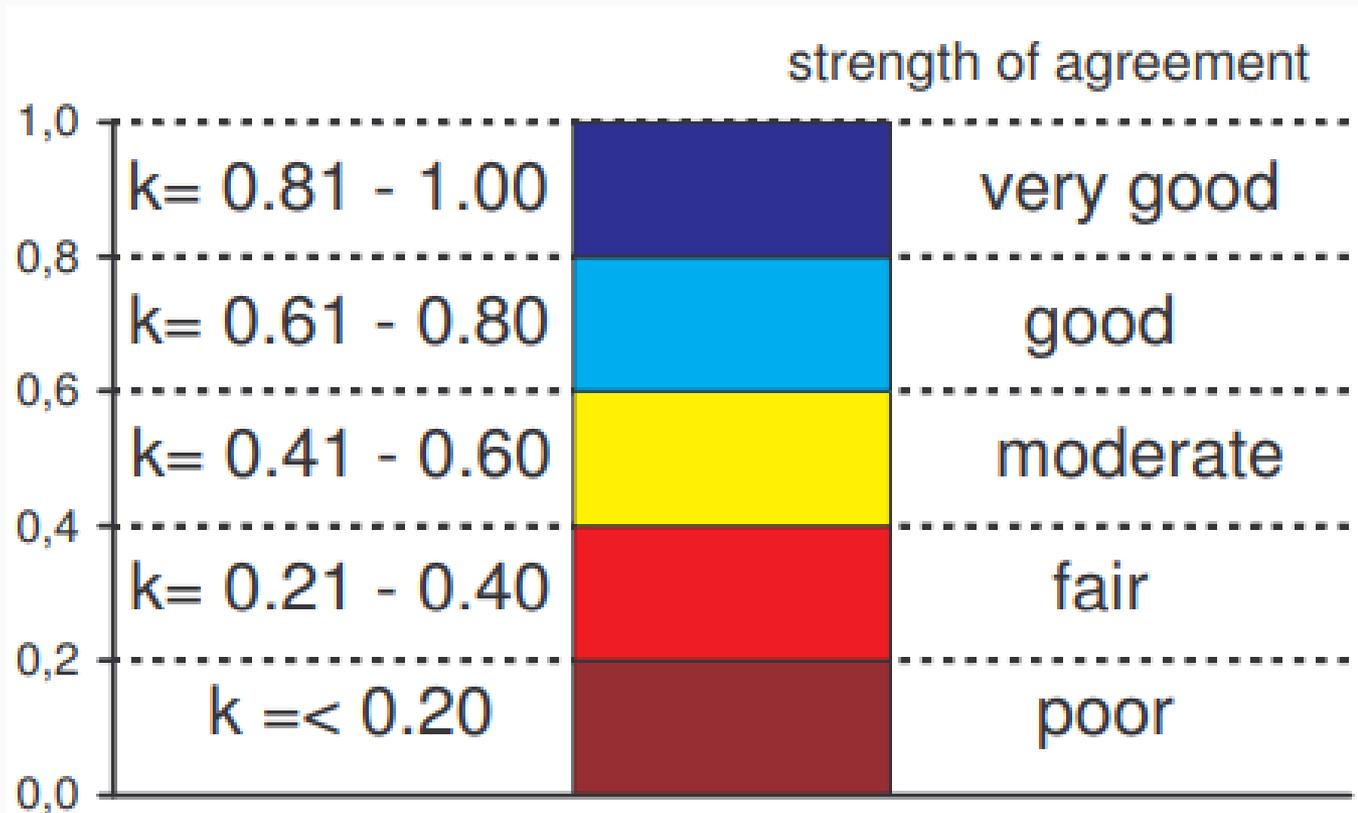
Accordo osservato 84%

Accordo dovuto a concordanza vera $84-52= 32\%$

Kappa= $32/48=$ 67%

Kappa di Cohen

Calcolo di kappa



Altman DG (1991) Practical statistics for medical research. Chapman & Hall: London.

Kappa di Cohen

Calcolo di kappa

$$k = \frac{(AO - AA)}{(1 - AA)}$$

Test di significatività

H0: $k=0$

Statistica z

$$z = \frac{K}{\sqrt{Var(K)}}$$

Se la probabilità p è inferiore alla significatività scelta (es. 0,05), rifiutiamo l'ipotesi nulla

Kappa di Cohen

Si deve valutare la condordanza sulla diagnosi da parte di due valutatori.

Lo studio viene condotto selezionando un gruppo di soggetti rappresentativo della popolazione "target" per il tipo di quesito diagnostico (ad esempio nel nostro caso preparati citologici di escreato bronchiale prelevati su soggetti a rischio di tumore polmonare).

I vetrini sono poi sottoposti a due patologi per la diagnosi, in cieco.

La diagnosi e' espressa secondo le seguenti categorie:

1. Negativo
2. Displasia lieve
3. Displasia grave
4. Carcinoma
5. Non valutabile

Kappa di Cohen

Tavola di contingenza raterA * raterB

Conteggio

| | | raterB | | | | | Totale |
|--------|---|--------|----|----|----|----|--------|
| | | 1 | 2 | 3 | 4 | 5 | |
| raterA | 1 | 26 | 19 | 1 | 0 | 7 | 53 |
| | 2 | 2 | 11 | 5 | 3 | 4 | 25 |
| | 3 | 0 | 1 | 6 | 6 | 0 | 13 |
| | 4 | 0 | 0 | 0 | 4 | 1 | 5 |
| | 5 | 1 | 1 | 0 | 0 | 2 | 4 |
| Totale | | 29 | 32 | 12 | 13 | 14 | 100 |

Accordo
osservato=
 $26+11+6+4+2=49$

Tavola di contingenza raterA * raterB

Conteggio atteso

| | | raterB | | | | | Totale |
|--------|---|--------|------|------|------|------|--------|
| | | 1 | 2 | 3 | 4 | 5 | |
| raterA | 1 | 15,4 | 17,0 | 6,4 | 6,9 | 7,4 | 53,0 |
| | 2 | 7,3 | 8,0 | 3,0 | 3,3 | 3,5 | 25,0 |
| | 3 | 3,8 | 4,2 | 1,6 | 1,7 | 1,8 | 13,0 |
| | 4 | 1,5 | 1,6 | ,6 | ,7 | ,7 | 5,0 |
| | 5 | 1,2 | 1,3 | ,5 | ,5 | ,6 | 4,0 |
| Totale | | 29,0 | 32,0 | 12,0 | 13,0 | 14,0 | 100,0 |

Accordo casuale=
 $15,4+8+1,6+1,7+1,6=$
26,3

$k=(49-26,3)/(1-26,3)=$
0,31

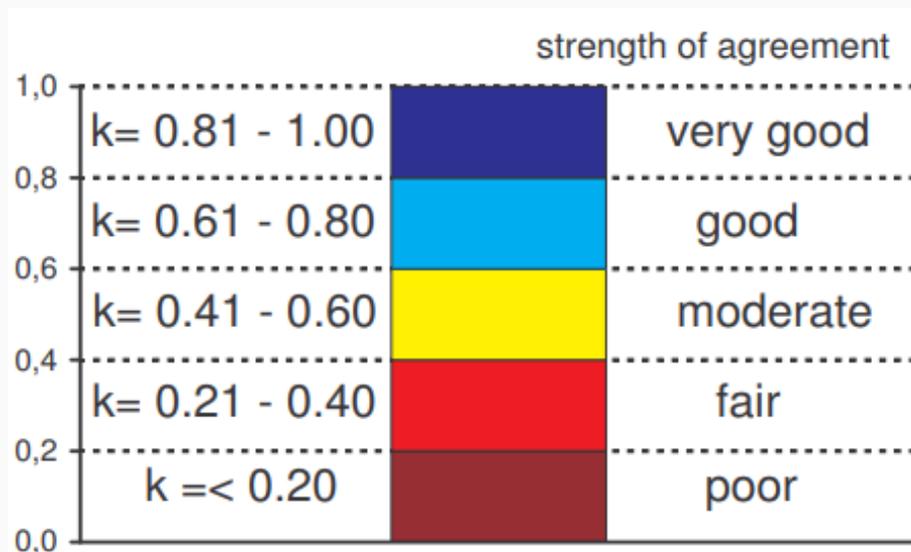
Kappa di Cohen

Misure simmetriche

| | Valore | E.S. asint. ^a | T appross. ^b | Sig. appross. |
|-----------------------------|--------|--------------------------|-------------------------|---------------|
| Misura di concordanza Kappa | ,310 | ,062 | 5,863 | ,000 |
| N. di casi validi | 100 | | | |

a. Senza assumere l'ipotesi nulla.

b. Viene usato l'errore standard asintotico in base all'assunzione dell'ipotesi nulla.



Accuratezza

La formulazione di una diagnosi è un processo imperfetto

Informazioni complesse (cliniche, strumentali...) vengono ridotte a semplici dicotomie necessarie per decidere se intraprendere o meno un determinato trattamento

Un modo semplificato di guardare alla relazione tra risultato di un test e la vera diagnosi:

| | | Malattia | |
|------|---|----------------|----------------|
| | | + | - |
| Test | + | Veri Positivi | Falsi positivi |
| | - | Falsi negativi | Veri negativi |

La valutazione dell'accuratezza di un test è in relazione alla conoscenza dello stato di malattia o più semplicemente al risultato di un test definito gold standard

Molto spesso il gold standard è un test relativamente elaborato, costoso o invasivo (es. biopsia, procedure radiologiche invasive etc.)

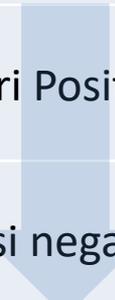
Un altro test spesso sostituisce il gold standard ed è usato come proxy per stabilire la presenza/assenza di malattia, con la consapevolezza che il rischio di misclassificazione può essere giustificato dalla fattibilità del test

Sensibilità

Proporzione di soggetti con la malattia che il test ha correttamente selezionato

Un test sensibile raramente misclassifica un soggetto con la malattia

| | | Malattia | |
|------|---|----------------|----------------|
| | | + | - |
| Test | + | Veri Positivi | Falsi positivi |
| | - | Falsi negativi | Veri negativi |



Sensibilità

Proporzione di soggetti con la malattia che il test ha correttamente selezionato

Un test sensibile raramente misclassifica un soggetto con la malattia

Diagnosi clinica vs coltura faringea

Faringite da Steptococco b emolitico gruppo A

| | | Malattia | |
|------|---|----------|----------------|
| | | + | - |
| Test | + | 27 | Falsi positivi |
| | - | 10 | Veri negativi |

Sensibilità=
 $27 / (27+10) = 73\%$

Specificità

Proporzione di soggetti senza malattia risultati negativi al test

Un test sensibile raramente misclassifica un soggetto con la malattia

Diagnosi clinica vs coltura faringea

Faringite da Steptococco b emolitico gruppo A

| | | Malattia | |
|------|---|----------------|----|
| | | + | - |
| Test | + | Veri positivi | 35 |
| | - | Falsi negativi | 77 |

Specificità=
 $77 / (77+35) = 69\%$

Sensibilità

In generale un test sensibile dovrebbe essere scelto quando la misclassificazione della presenza di malattia (falsi negativi) può arrecare grave danno al paziente

Es. se sospettiamo una malattia grave ma curabile (es. linfoma di Hodgkin, tubercolosi)

Un test sensibile è utile anche negli stadi precoci del processo diagnostico per ridurre il numero di diagnosi possibili

Un test sensibile si dice è usato «to rule out» disease

Specificità

Test specifici sono usati per confermare la malattia («to rule in»)

Un test altamente specifico raramente sarà negativo in presenza di malattia

Sono particolarmente utili quando un falso positivo può esporre il paziente a un danno alla salute o trauma emotivo

Es. quando abbiamo il sospetto di una diagnosi di tumore maligno, prima di esporre il paziente al trattamento è necessario confermare la diagnosi attraverso un esame molto specifico (es. biopsia)

Valori predittivi

Nella pratica clinica però quando è disponibile il risultato di un test, la sensibilità e la specificità non sono più molto utili poiché questi valori indicano la probabilità che il test sia positivo o negativo in persone di cui conosciamo lo stato di salute (presenza / assenza di malattia)

Per il clinico il grande problema è la determinazione della presenza di malattia dato il risultato del test

Valore predittivo del test: probabilità di malattia o di non malattia dato il risultato

Valore predittivo positivo:

probabilità di malattia dato il risultato positivo del test

Valore predittivo positivo: $\text{Veri positivi} / (\text{Veri positivi} + \text{Falsi positivi})$

| | | Malattia | |
|------|---|----------------|----------------|
| | | + | - |
| Test | + | Veri Positivi | Falsi positivi |
| | - | Falsi negativi | Veri negativi |

Valore predittivo negativo:

probabilità di non malattia dato il risultato negativo del test

Valore predittivo negativo:

Veri negativi / (Veri negativi + Falsi negativi)

| | | Malattia | |
|------|---|----------------|----------------|
| | | + | - |
| Test | + | Veri Positivi | Falsi positivi |
| | - | Falsi negativi | Veri negativi |

Valore predittivo positivo

Diagnosi clinica vs coltura faringea

Faringite da Steptococco b emolitico gruppo A

| | | Malattia | |
|------|---|----------|----|
| | | + | - |
| Test | + | 27 | 35 |
| | - | 10 | 77 |

Valore predittivo positivo: probabilità di malattia dato il risultato positivo del test

Valore predittivo positivo: Veri positivi / (Veri positivi + Falsi positivi)

Valore predittivo positivo: $27 / (27 + 35) = 0,44$; **44%**

Valore predittivo negativo

Diagnosi clinica vs coltura faringea

Faringite da Steptococco b emolitico gruppo A

| | | Malattia | |
|------|---|----------|----|
| | | + | - |
| Test | + | 27 | 35 |
| | - | 10 | 77 |

Valore predittivo negativo: probabilità di non malattia dato il risultato negativo del test

Valore predittivo negativo: Veri negativi / (Veri negativi + Falsi negativi)

Valore predittivo negativo: $77 / (77 + 10) = 0,89$ **89%**

Valore predittivo e prevalenza

La sensibilità e la specificità sono proprietà intrinseche del test. Il valore predittivo di un test è determinato dalla sensibilità e dalla specificità di un test, ma anche dalla prevalenza della malattia nella popolazione nella quale viene sperimentato

Bassa prevalenza : aumentano i falsi positivi

Alta prevalenza: aumentano i falsi negativi

$$PPV = \frac{(Sensitivity)(p)}{(Sensitivity)(p) + (1 - Specificity)(1 - p)}$$

$$NPV = \frac{(Specificity)(1 - p)}{(Specificity)(1 - p) + (1 - Sensitivity)(p)}$$

p = prevalence of the disease in the population being studied
expressed as a fraction (ex. 0.20 for 20% prevalence)

L'interpretazione di un test varia da un setting assistenziale ad un altro in accordo con la prevalenza della patologia nella specifica popolazione

Valore predittivo e prevalenza

Mammografia per la diagnosi di Carcinoma Mammario

(Se=75%; Sp=92.3%)

Prevalenza Ca mammario in età 45-74

| | | Ca mammario | | Tot |
|-----|-----|-------------|------|-------|
| | | Si | No | |
| Mx | Pos | 41 | 796 | 837 |
| | Neg | 13 | 9150 | 9163 |
| Tot | | 54 | 9946 | 10000 |

$$\text{VPP} = 41/837 = 4.9\%$$

$$\text{VPN} = 9150/9163 = 99.9\%$$

Prevalenza Ca mammario in età 65-69

| | | Ca mammario | | Tot |
|-----|-----|-------------|------|-------|
| | | Si | No | |
| Mx | Pos | 88 | 791 | 879 |
| | Neg | 29 | 9092 | 9121 |
| Tot | | 117 | 9883 | 10000 |

$$\text{VPP} = 88/879 = 10.0\%$$

$$\text{VPN} = 9092/9121 = 99.7\%$$

Rapporto di verosimiglianza

Possiamo descrivere la performance di un test in un modo sintetico che tiene conto contemporaneamente di sensibilità e specificità: il rapporto di verosimiglianza (likelihood ratio LR)

Possiamo calcolare un **LR positivo** e un **LR negativo** e rispondono alle domande, rispettivamente:

- un risultato positivo al test, quante volte è più verosimile trovarlo nella pop. dei malati rispetto a quella dei non malati?
- un risultato negativo al test, quante volte è più verosimile trovarlo nella pop. dei malati rispetto a quella dei non malati?

$$LR^+ = \frac{\textit{sensitivity}}{1 - \textit{specificity}}$$

$$LR^- = \frac{1 - \textit{sensitivity}}{\textit{specificity}}$$

Rapporto di verosimiglianza

Diagnosi clinica vs coltura faringea

Faringite da Steptococco b emolitico gruppo A

| | | Malattia | |
|------|---|----------|----|
| | | + | - |
| Test | + | 27 | 35 |
| | - | 10 | 77 |

$$LR^+ = \frac{\textit{sensitivity}}{1 - \textit{specificity}}$$

$$LR^- = \frac{1 - \textit{sensitivity}}{\textit{specificity}}$$

$$LR^+ = 0,73 / (1 - 0,69) = \mathbf{2,35}$$

$$LR^- = (1 - 0,73) / 0,69 = \mathbf{0,39}$$

Interpretazione del rapporto di verosimiglianza

| | |
|--------------------|---------------------------------|
| LR+ >10 : | eccellente accuratezza del test |
| LR+ tra 5 e 10: | buona accuratezza del test |
| LR+ tra 2 e 5: | mediocre accuratezza del test |
| LR+ <2: | test inutile |
| LR- tra 0,5 e 1: | test inutile |
| LR- tra 0,2 e 0,5: | mediocre accuratezza del test |
| LR- tra 0,1 e 0,2 | buona accuratezza del test |
| LR- <0,1: | eccellente accuratezza del test |

Relazione sensibilità / specificità

Scelta del cut-off tra valori normali e patologici

Per un risultato del test espresso su una scala di valori, possiamo aumentare la sensibilità (o la specificità) solo a spese della specificità (o sensibilità)

Trade-Off between Sensitivity and Specificity when Diagnosing Diabetes^a

| BLOOD SUGAR LEVEL 2 HOURS AFTER EATING | SENSITIVITY | SPECIFICITY |
|---|-------------|-------------|
| <i>mg/100 ml</i> | <i>%</i> | <i>%</i> |
| 70 | 98.6 | 8.8 |
| 80 | 97.1 | 25.5 |
| 90 | 94.3 | 47.6 |
| 100 | 88.6 | 69.8 |
| 110 | 85.7 | 84.1 |
| 120 | 71.4 | 92.5 |
| 130 | 64.3 | 96.9 |
| 140 | 57.1 | 99.4 |
| 150 | 50.0 | 99.6 |
| 160 | 47.1 | 99.8 |
| 170 | 42.9 | 100.0 |
| 180 | 38.6 | 100.0 |
| 190 | 34.3 | 100.0 |
| 200 | 27.1 | 100.0 |

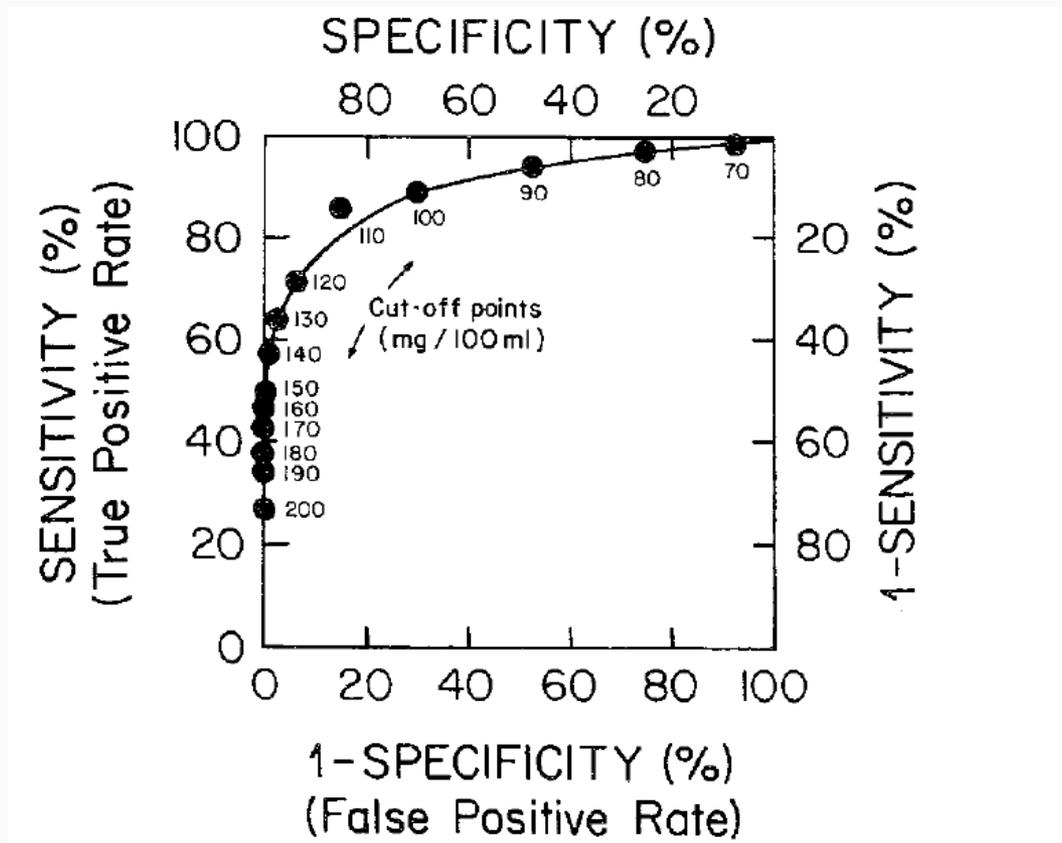
^a From *Diabetes Program Guide*, Public Health Service Publication No. 506, 1960.

Relazione sensibilità / specificità

Un altro modo per esprimere la relazione sensibilità / specificità per un test è di costruire una curva, chiamata receiver operator characteristic (ROC)

La curva ROC è utilizzata per descrivere l'accuratezza di un test diagnostico su un range di cutoff

L'accuratezza del test viene sintetizzata da una misura, l'area sotto la curva



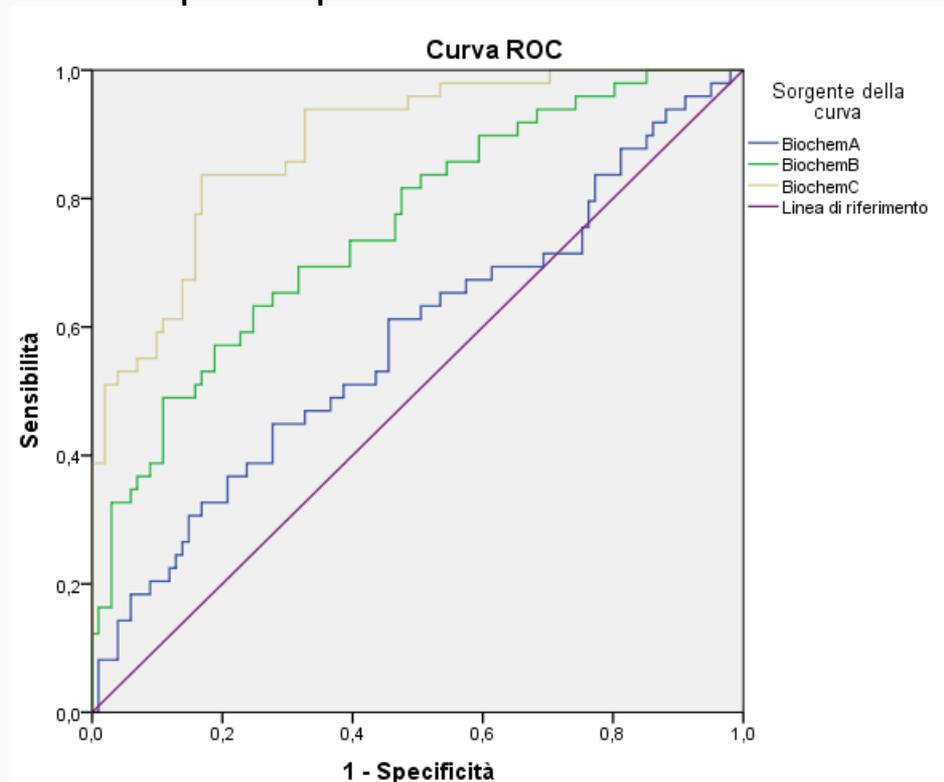
Curva ROC

Esempio

Analizziamo i dati di tre diversi test biochimici. Un test genetico è stato identificato come gold standard per determinare la presenza / assenza di malattia

Le curve ROC sono usate per valutare quale test è migliore nel predire la malattia

H0: area sotto la curva (AUC) = 0,5 (la capacità del test di predire la malattia è pari a quella di una moneta)



Area sottostante la curva

| Variabili del risultato del test | Area | Errore std ^a | Sig. asintotica ^b | Intervallo di confidenza al 95% asintotico | |
|----------------------------------|------|-------------------------|------------------------------|--|------------------|
| | | | | Limite inferiore | Limite superiore |
| BiochemA | ,580 | ,051 | ,114 | ,479 | ,681 |
| BiochemB | ,755 | ,042 | ,000 | ,673 | ,837 |
| BiochemC | ,886 | ,028 | ,000 | ,832 | ,940 |

a. In base all'assunzione non parametrica

b. Ipotesi nulla: area reale = 0.5

Come calcolare il cut-off?

La scelta del cut-off può variare in funzione dell'obiettivo clinico

Es.

In un test di screening scelgo un cut-off che massimizza la sensibilità

In un test di conferma scelgo un cut-off che massimizza la specificità

In alternativa posso scegliere un cut-off che massimizza sensibilità e specificità

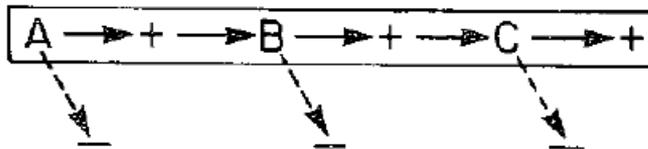
Test multipli

Due principali strategie diagnostiche:

| <u>STRATEGY</u> | <u>SEQUENCE OF EVENTS</u> | <u>CONSEQUENCES</u> |
|-----------------|---------------------------|---------------------|
|-----------------|---------------------------|---------------------|

- SERIAL TESTING

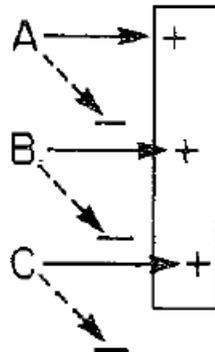
Test A *and* test B *and* test C are positive



↓ SENSITIVITY
↑ SPECIFICITY

- PARALLEL TESTING

Test A *or* test B *or* test C is positive



↑ SENSITIVITY
↓ SPECIFICITY