

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE
**CALCOLO DEL CUT-OFF NEL CASO DI TEST DIAGNOSTICI
CONTINUI**

Relatore Prof.ssa Laura Ventura
Dipartimento di Scienze Statistiche

Laureando: Francesca Di Vittorio
Matricola N. 1233220

Anno Accademico 2021/2022

*Alla mia famiglia e ai miei amici,
che mi hanno supportata lungo questo percorso*

Indice

Capitolo 1: I test diagnostici	4
1.1. Il potere diagnostico di un test.....	5
1.2. La Curva ROC.....	10
1.2.1. Costruzione della Curva ROC.....	11
1.2.2. Area sottesa alla Curva ROC (AUC).....	14
1.2.3. Confronto fra curve ROC.....	16
Capitolo 2: Criteri di selezione del Cut-Off	18
2.1. L'indice J di Youden.....	21
2.2. Metodo del punto di simmetria (SP=SE).....	23
2.2.1. Inferenza parametrica (GPQ).....	25
2.2.2. Inferenza non parametrica (EL).....	26
2.2.3. Studio di una simulazione.....	30
Capitolo 3: Applicazione dei metodi per la ricerca del cut-off	32
3.1. Esempio: Leishmaniosi canina.....	33
3.2. Esempio: Linfomi.....	38
Conclusioni	42
Bibliografia	43

CAPITOLO 1

I test diagnostici

I test diagnostici sono esami il cui scopo è diagnosticare se un individuo abbia o meno una certa patologia. Sono strumenti clinici importanti per la diagnosi delle malattie, che è uno dei tre processi decisionali della medicina clinica insieme alla prognosi e alla terapia. In una situazione ideale, ci si aspetta che un test diagnostico discrimini perfettamente i malati dai sani, ossia, una volta eseguito il test, gli individui vengono classificati con assoluta certezza come affetti o non affetti dalla malattia di interesse. In questo caso si parla di un test “gold standard o standard di riferimento” che però non esiste o è molto difficile o costoso da eseguire per determinate condizioni patologiche.

Nella realtà, un determinato test diagnostico non discrimina in maniera esatta i malati dai sani, ossia il test può identificare risultati:

- veri positivi (TP), cioè la proporzione di individui che hanno un valore alterato del test tra tutti quelli realmente affetti dalla malattia;
- veri negativi (TN), cioè la proporzione di individui che non hanno un valore alterato del test e sono sani;
- falsi positivi (FP), cioè la proporzione di individui che pur avendo un valore alterato del test non sono affetti dalla malattia di interesse;
- falsi negativi (FN), cioè la proporzione di individui che pur non avendo un valore alterato del test sono malati.

I quattro valori identificati (TP, FP, TN, FN) possono essere rappresentati in una tabella a doppia entrata che conta il numero di unità classificate correttamente o meno per ciascuna delle due modalità possibili. Tale matrice è detta matrice di confusione (o tabella di errata classificazione) (Tabella 1).

Risultato del test	Malato D ⁺	Sano D ⁻
Positivo T ⁺	TP	FN
Negativo T ⁻	FP	TN

Tabella 1: Matrice di confusione

I valori sulla diagonale rappresentano la proporzione di unità correttamente classificate, gli altri sono gli errori.

1.1 Il potere diagnostico di un test

Per l'interpretazione del risultato di un esame è necessario conoscere alcune caratteristiche del test, quali la sensibilità e la specificità diagnostica in riferimento alla malattia nella popolazione studiata, per poter valutare il valore predittivo che ci indica quale percentuale di pazienti con risultati positivi è da considerarsi ammalata, almeno dal punto di vista statistico. Per comprendere il significato di questi indici si considera la Tabella 1 che mette in rapporto, in termini generali, i risultati di un ipotetico test diagnostico (che può essere positivo o negativo) con la presenza/assenza di una specifica malattia.

La sensibilità (Se) di un test è impiegata per indicare la proporzione di risposte positive che si ottengono applicando l'esame a pazienti affetti dalla malattia, ovvero la probabilità che una persona malata risulti positiva al test:

$$Se = \frac{TP}{TP+FP} .$$

La specificità del test (Sp), invece, è impiegata per indicare la proporzione delle risposte negative che si ottengono applicando l'esame a pazienti sani, ovvero la probabilità che una persona sana risulti negativa al test:

$$Sp = \frac{TN}{TN+FN} .$$

Il potere predittivo positivo è la probabilità a posteriori della presenza di una malattia quando ci si trova a interpretare un risultato positivo di un esame. È la proporzione di pazienti malati tra tutti quelli che sono positivi al test:

$$VP+ = \frac{TP}{TP+FN} .$$

Il potere predittivo negativo rappresenta la probabilità che un soggetto sano con un esame negativo non abbia effettivamente la malattia. È la proporzione di pazienti sani tra tutti quelli che sono negativi al test:

$$VP- = \frac{TN}{TN+FP} .$$

L'accuratezza è la proporzione di pazienti correttamente classificati, ed è il rapporto tra la somma delle persone classificate correttamente quindi i veri positivi (TP) e i veri negativi (TN) sul totale delle persone studiate ($TOT=TP+FP+TN+FN$):

$$Accuratezza = \frac{TP+TN}{TOT} .$$

La sensibilità e la specificità sono misure indipendenti dalla prevalenza della malattia, cioè non sono influenzate dalla frequenza con cui una certa patologia è presente in uno specifico campione di individui. Viceversa, il potere predittivo negativo e positivo sono invece strettamente dipendenti dalla frequenza della malattia di interesse.

Si dice che un test è sensibile al 100% quando tutti i malati sono risultati positivi, mentre è specifico al 100% quando tutti i sani risultano negativi.

Se il risultato del test diagnostico di interesse è una variabile binaria (affetto/non affetto), per discriminare i pazienti nei due gruppi è sufficiente calcolare la sensibilità, la specificità, il potere predittivo positivo, il potere predittivo negativo e l'accuratezza. Ad esempio, supponiamo che il test sia per la discriminazione di due soli stati, "malato" (D^+) e "sano" (D^-) per la discriminazione dei soggetti nei due gruppi, si deve impostare un valore di cut-off; i risultati del test uguali o superiori a questo valore sono considerati positivi (T^+), in caso contrario sono negativi (T^-). Ci sono, tre situazioni che si ottengono spesso nell'interpretazione del test:

1. La prima è il caso di test ideale che discrimina con precisione assoluta, i pazienti sani D^- dai malati D^+ come mostrato in Figura 1.

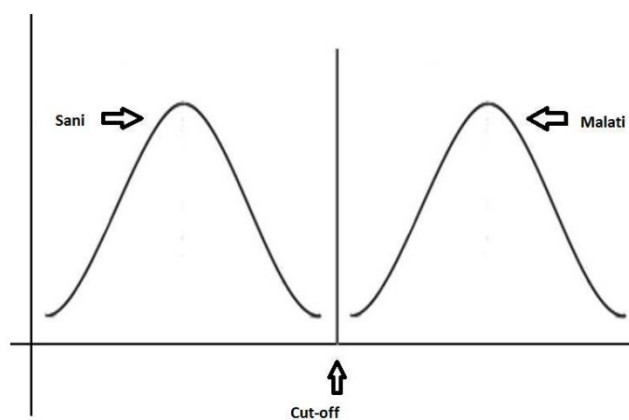


Figura 1

2. Nell'altra, invece, il caso è opposto e non si riesce ad attribuire il paziente al gruppo di sani o malati. In questo caso il test non ha potere diagnostico (Figura 2).

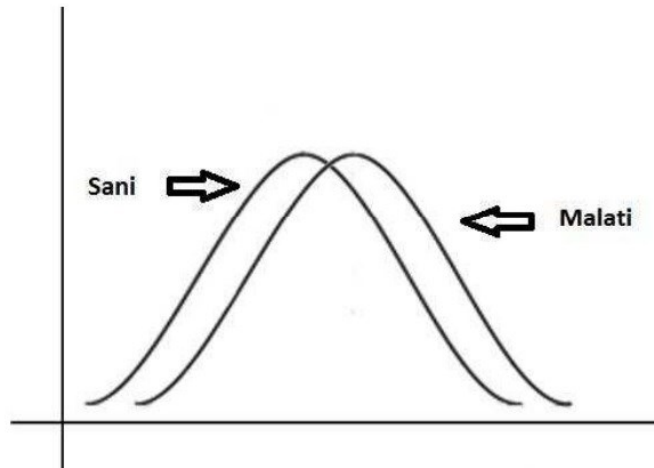


Figura 2

3. Infine, il caso più frequente è quello mostrato in Figura 3, in cui ci sono zone più o meno ampie di sovrapposizione delle distribuzioni dei pazienti malati e sani.

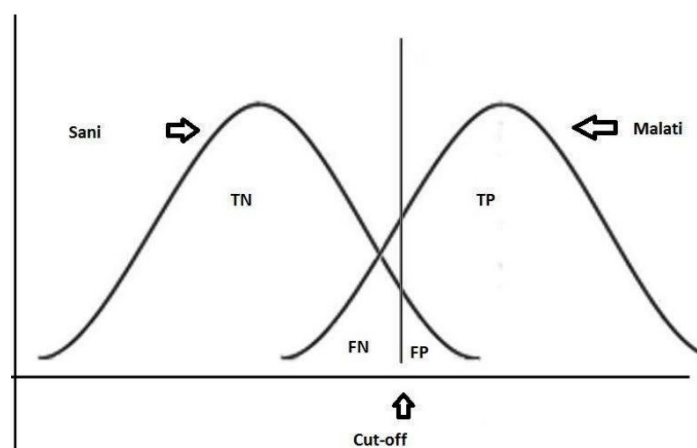


Figura 3

La scelta del valore di cut-off, quindi, determina le percentuali di risultati dei test veri positivi (TP), veri negativi (TN), falsi positivi (FP) e falsi negativi (FN).

Pertanto, un cut-off elevato permetterà di identificare correttamente la maggior parte dei sani, conferendo al test un'elevata specificità (quindi pochi FP), ma sottostimerà la proporzione dei malati, conferendo al test una bassa sensibilità (quindi molti FN). Mentre un cut-off basso, al contrario, permetterà di identificare correttamente la maggior parte dei malati, conferendo al test un'elevata sensibilità (pochi FN), ma sottostimerà la proporzione dei sani, conferendo al test una bassa specificità (molti FP). Pertanto, è necessario specificare la scelta del livello discriminante che può variare a seconda della situazione clinica in cui l'indagine viene utilizzata. In particolare, quando una malattia è grave e non si dovrebbe correre il rischio di non diagnosticarla, oppure quando la malattia è curabile e i risultati falsi positivi non comportino danni né di tipo psicologico né di tipo economico per il paziente, verrà scelto un cut-off caratterizzato da massima sensibilità clinica. Viceversa, verrà scelto un valore dotato di massima specificità clinica se la malattia, pur essendo grave, non è curabile, oppure quando escludere la malattia abbia un elevato valore psicologico e i risultati falsi positivi possono comportare gravi danni per il paziente. Per esempio, negli esami di screening in cui si vuole identificare tutti i soggetti affetti da una malattia, occorre disporre di un'indagine estremamente sensibile accettando una bassa specificità; in tal modo aumenta il numero di soggetti che pur essendo sani manifestano positività del risultato, ma si riduce al minimo la possibilità di non conoscere soggetti affetti dalla malattia. Una successiva indagine di conferma rende quindi necessaria l'identificazione di falsi positivi che devono avere un'alta specificità, allo scopo di identificare i falsi positivi.

1.2 La Curva ROC

La sensibilità e la specificità, il potere predittivo negativo e positivo classificano gli individui come affetti o non affetti da una specifica malattia sulla base di un predefinito valore del test (valore soglia), quindi si riferiscono a un risultato binario. Il dato in laboratorio, invece, è per natura quantitativo, ovvero molto frequentemente i test hanno risultati continui; quindi, bisogna adottare un metodo alternativo per discriminare i malati dai sani senza perdere informazioni che possono essere anche molto rilevanti. Infatti l'interpretazione clinica dipende molto dal grado di positività e negatività del risultato. Inoltre, l'uso del valore predittivo risulta valido solo per un'analisi univariata e non considera l'incertezza associata al risultato di un'indagine in laboratorio. Per questo, il metodo più opportuno per misurare l'accuratezza di un test diagnostico lungo tutto il range dei valori possibili è l'uso della Curva ROC (*Receiver Operating Characteristic*) (Figura 4).

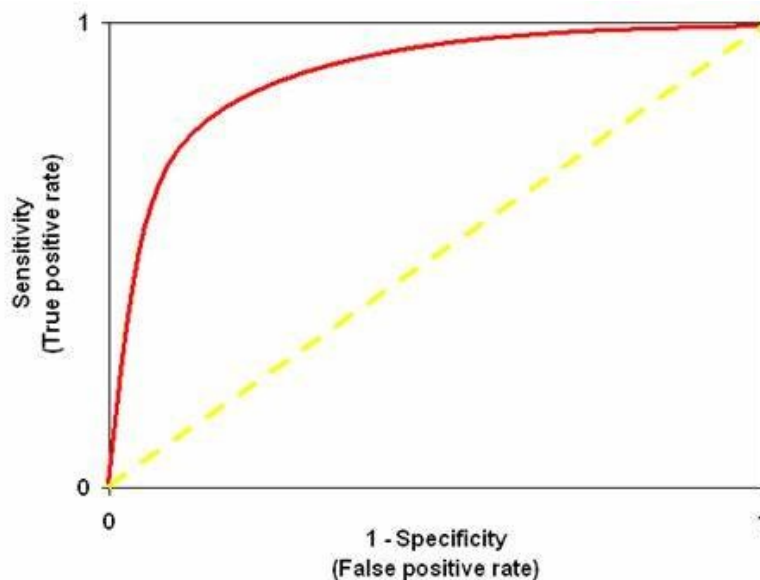


Figura 4: Rappresentazione della curva ROC

L'uso di questo metodo risale alla Seconda guerra mondiale, quando la capacità degli operatori radar (ricevitori) è stata testata per determinare se un "blip" sullo

schermo del radar rappresentasse un oggetto (segnale, un risultato TP) o un rumore (un risultato FP). Diversi anni dopo, il metodo è stato ritenuto utile in molte altre discipline scientifiche, inclusa la medicina diagnostica, in cui un medico dovrebbe discriminare un risultato del test veramente positivo (TP) da un falso positivo (FP). Recentemente è entrata anche nell'ambito del data mining (Metz, 1978; Pepe, 2003; Zou, 2002).

1.2.1 Costruzione della Curva ROC

Per costruire la curva ROC si considerano tutti i possibili cut-off e, per ognuno di questi, si calcola la proporzione di veri positivi (la sensibilità) e la proporzione di falsi positivi. La proporzione di falsi positivi si calcola con la formula standard: $1 - \text{specificità}$.

Congiungendo i punti che mettono in rapporto la proporzione di veri positivi e di falsi positivi, si ottiene la curva ROC. Questa tecnica statistica misura l'accuratezza di un test tramite la misura dell'area sottostante alla curva ROC (AUC, *Area Under the Curve*).

Tale curva offre un'illustrazione grafica della summenzionata relazione tra la sensibilità e specificità di un test e rappresenta il tasso TP (Se) rispetto al tasso FP ($1 - Sp$) per ciascun valore di cut-off.

La struttura generale di una curva ROC è semplice. La curva è confinata in un quadrato unitario (Figura 5). Nell'angolo inferiore sinistro la specificità è uguale a 0 e la sensibilità è uguale ad 1, al variare del valore di cut-off, la sensibilità aumenta e la specificità del test diminuisce, spostandosi dalla curva dall'angolo inferiore sinistro fino all'angolo superiore destro del quadrato dove $Se = 1$ e $Sp = 0$.

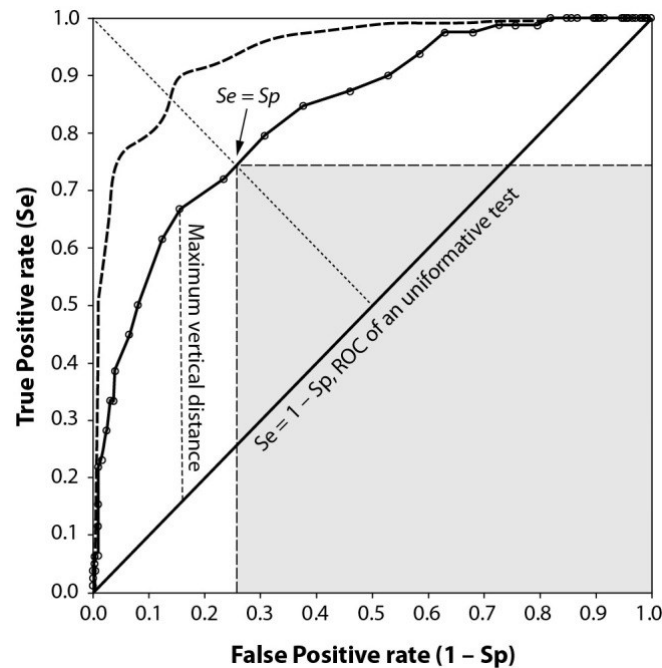


Figura 5: La struttura generale di una curva ROC. La curva (linea tratteggiata) che si trova completamente al di sopra di un'altra curva (linea continua), rappresenta un test migliore. L'area della regione rettangolare ombreggiata è massima quando i suoi lati (Se e Sp) sono uguali.

In un test perfetto, sia Se che Sp sono uguali a 1. La curva ROC corrispondente a un test perfetto (cioè il gold standard) è un segmento di linea che collega l'angolo inferiore sinistro all'angolo superiore sinistro e quello destro all'angolo superiore sinistro (una curva che coincide con i lati sinistro e superiore del quadrato). D'altra parte, la curva ROC corrispondente a un test senza valore diagnostico è il segmento di linea che collega l'angolo inferiore sinistro all'angolo superiore destro, ovvero la bisettrice (*"chance line"*) (Bamber, 1975; Zweig, Campbell, 1993). La curva ROC solitamente si trova tra questi due estremi (Figura 6). Siano X e Y variabili casuali con distribuzione F_x e F_y che rappresentano rispettivamente la popolazione sana e malata. Allora la curva *ROC* è data da

$$ROC(p) = 1 - F_x((F_y)^{-1}(1-p)),$$

con $p \in (0,1)$, $p = 1 - F_Y(\tau)$, con τ cut-off. Se ad esempio le variabili hanno distribuzione normale, ossia $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$, la curva ROC è data da

$$ROC(p) = \Phi \left(\frac{\mu_y - \mu_x}{\sigma_y} + \sigma_x \frac{\Phi^{-1}(p)}{\sigma_y} \right),$$

con Φ funzione di ripartizione di una normale standard.

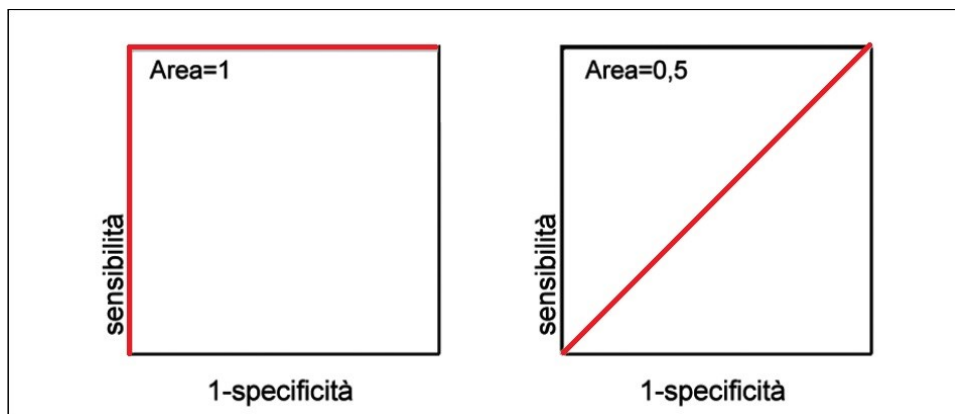


Figura 6: Esempio di test perfettamente discriminante (pannello di sinistra) e di test che non discrimina i malati dai sani (pannello di destra).

1.2.2. Area sottesa alla curva ROC

Matematicamente, l'area sotto la curva ROC (AUC), è equivalente alla probabilità che il risultato del test misurato in una persona malata (D^+) selezionata casualmente sia maggiore di quello misurato in una persona sana (D^-). Il calcolo dell'AUC per una curva empirica (Figura 5) può essere fatta connettendo i diversi punti della curva ROC all'asse delle ascisse con segmenti verticali e sommando le aree dei risultanti poligoni generati nella zona sottostante. Tale tecnica è chiamata regola trapezoidale, ma potrebbe fornire risultati distorti. Una seconda possibilità utilizza un modello di sollecitazione-resistenza (stress-strength model) (vedi, ad esempio, Kotz *et al.*, 2003) che esprime l'AUC come la probabilità che i pazienti sani (X) siano minori dei pazienti malati (Y):

$$AUC = P(X < Y).$$

Con questo modello si possono fare assunzioni parametriche e non parametriche sulle variabili X e Y .

Sotto ipotesi non parametriche, la quantità AUC è collegata al test U di Mann-Whitney (Bottarelli e Parodi, 2003). Tale test viene utilizzato per confrontare la distribuzione di una variabile continua tra due gruppi e per testare che i due gruppi abbiano la stessa mediana. Tale ipotesi è uguale a testare che un soggetto preso a caso dal gruppo X abbia la stessa probabilità di avere un valore della variabile inferiore ad un valore predefinito di quello di un soggetto estratto a caso dal gruppo Y (Bottarelli e Parodi, 2003). La statistica è:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1,$$

con n_1 e n_2 numerosità campionarie dei due gruppi e R_1 somma dei ranghi del gruppo con numerosità n_1 . Da qui Bamber (1975) ha dimostrato che l'area sottesa alla curva ROC, sotto ipotesi non parametriche, si può calcolare con la seguente formula:

$$AUC = 1 - \frac{U}{n_1 n_2}.$$

Sotto ipotesi parametriche, invece, si assume in genere che X e Y siano variabili casuali con funzione di densità, $f_x(x; \theta_x)$ e $f_y(y; \theta_y)$, con $\theta_x \in \Theta \subseteq R$ e $\theta_y \in \Theta \subseteq R$. L'AUC può essere espressa come

$$AUC = AUC(\Theta) = P(X < Y) = \int_{-\infty}^{+\infty} F_x(t; \theta_x) dF_y(t; \theta_y),$$

con F_x e F_y funzioni di ripartizione di X e Y , rispettivamente.

Esistono diverse espressioni dell'AUC a seconda della distribuzione delle due variabili. Nel caso più comune X e Y hanno distribuzione normale, ossia $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ e si ottiene

$$AUC = AUC(\theta) = P(X < Y) = \Phi\left(-\frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}\right),$$

con $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$.

Oppure se $X \sim \text{Exp}(\alpha)$ e $Y \sim \text{Exp}(\beta)$, si ottiene

$$AUC = AUC(\theta) = P(X < Y) = \frac{\alpha}{\alpha + \beta},$$

con $\theta = (\alpha, \beta)$.

Esiste un metodo parametrico che usa la proprietà di equivarianza degli stimatori di massima verosimiglianza (cfr. ad esempio Pace, Salvan, 2001, Cap 3), per cui l'AUC può essere stimato come

$$\widehat{AUC} = AUC(\hat{\theta}) = AUC(\hat{\theta}_x, \hat{\theta}_y).$$

Il valore dell'AUC varia tra 0.5 nel caso in cui il test non discriminasse per niente i malati dai sani, che coincide con l'area sottostante la bisettrice del grafico (Figura 6, pannello di destra) e 1 per un test perfetto che discrimina perfettamente i malati dai sani, cioè il 100% di accuratezza (Figura 6, pannello di sinistra).

Nella realtà si considera adeguato un test diagnostico con un'area sotto la curva $\geq 80\%$. Tanto maggiore è l'area sotto la curva (cioè tanto più la curva si avvicina al vertice del grafico) tanto maggiore è il potere discriminante del test (Figura 7).

Per l'interpretazione dei valori dell'area sottostante la curva ROC è possibile riferirsi alla classificazione proposta da Swets (1998):

- 1) $AUC=0.5$, il test non è informativo;
- 2) $0.5 < AUC \leq 0.7$, il test è poco accurato;
- 3) $0.7 < AUC \leq 0.9$, il test è moderatamente accurato;
- 4) $0.9 < AUC < 1$, il test è altamente accurato;
- 5) $AUC=1$, il test è perfetto.

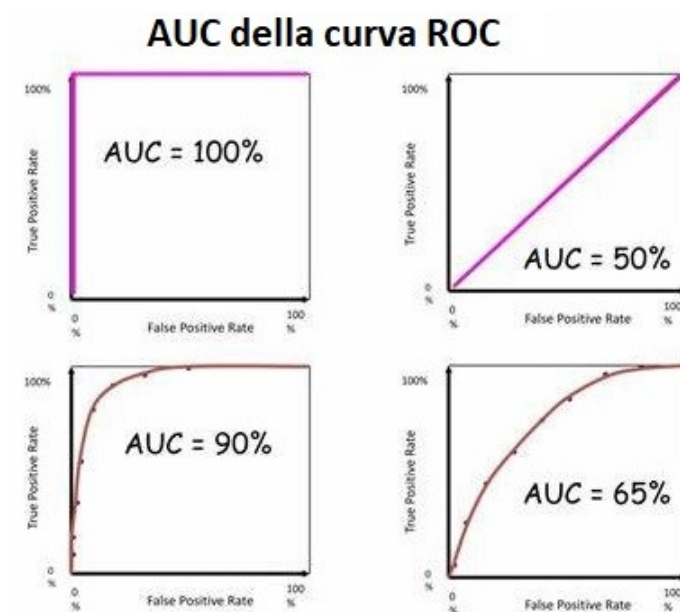


Figura 7: Esempi di rappresentazione dell'AUC in caso di test perfetto, test non informativo, test moderatamente accurato e test poco accurato.

1.3.3 Confronto fra curve ROC

La curva ROC può anche essere utilizzata per confrontare l'efficienza di esami differenti in una particolare patologia. Il confronto fra due esami può essere effettuato visivamente; la curva più spostata nella parte centrale del grafico appartiene all'esame peggiore (Figura 5), oppure anche in maniera oggettiva confrontando i due valori dell'AUC e preferendo quello maggiore. Esiste anche un

test (Hanley e McNeil, 1983) per confrontare due curve ROC indipendenti ed è dato da:

$$Z = \frac{\widehat{AUC}_1 - \widehat{AUC}_2}{\sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}} \stackrel{H_0}{\sim} N(0,1),$$

dove $\widehat{\sigma}_1^2$ e $\widehat{\sigma}_2^2$ sono le varianze di \widehat{AUC}_1 e \widehat{AUC}_2 , rispettivamente.

Se i due test non sono indipendenti, ovvero quando vengono applicati agli stessi soggetti, bisogna aggiungere la correlazione dell'errore standard della differenza delle due aree e diventa (Hanley e McNeil, 1982):

$$Z = \frac{\widehat{AUC}_1 - \widehat{AUC}_2}{\sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2 - 2r\widehat{\sigma}_1\widehat{\sigma}_2}} \stackrel{H_0}{\sim} N(0,1)$$

Capitolo 2

Criteri di selezione del cut-off

Nella ricerca medica, i test diagnostici con valori continui sono ampiamente impiegati per la discriminazione di soggetti malati e non malati. L'accuratezza diagnostica di un test può essere valutata utilizzando la curva ROC che rappresenta graficamente i vari livelli discriminanti di un esame diagnostico e permette di identificare il valore soglia ottimale, il cosiddetto "best cut-off", che è quel valore del test che minimizza gli errori di classificazione. Quando la distribuzione dei valori nelle due classi malati-sani è di tipo normale, la soglia ottimale è pari al valore di un'ascissa corrispondente al punto di intersezione delle distribuzioni dei due gruppi (Bottarelli, Parodi, 2003) (Figura 8).

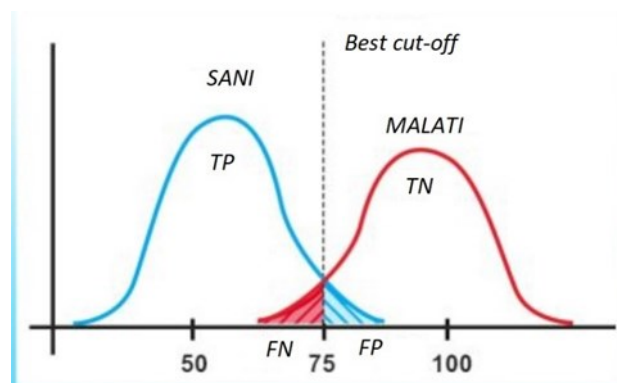


Figura 8: Rappresentazione del "best cut-off" nel punto di intersezione delle due distribuzioni

Tuttavia, la scelta del cut-off non può essere dettata solo da considerazioni di ordine probabilistico volte a minimizzare la proporzione di classificazioni errate, ma è necessario basarsi anche sul prevedibile impatto di tipo sanitario, economico, sociale, ecc. di ciascuno dei due tipi di mis-classificazione (FP e FN). Ad esempio per malattie di alta contagiosità è opportuno minimizzare la soglia di falsi negativi.

Quindi per determinare la soglia ottimale, “cut-off” possono essere considerati diversi approcci, come quelli basati sul cosiddetto punto vicino a (0,1), sull'indice di Youden, e sul punto di simmetria.

Un metodo empirico utilizzato per la scelta del cut-off consiste nel fissare a priori il valore desiderato di specificità (generalmente $>0,9$) e quindi calcola la corrispondente sensibilità del test nella suddetta condizione. Questo approccio genera tuttavia due effetti collaterali negativi; non permette di trovare il cut-off ottimale, in quanto il test potrebbe produrre risultati migliori con altri valori e non si riescono a confrontare le performance di due o più test valutati in base ad un singolo valore di cut-off. Pertanto vengono ostacolati gli studi di meta-analisi nei quali vengono ottenute comparazioni qualitative fra risultati ottenuti in studi diversi sullo stesso argomento.

Nella maggior parte degli studi, si assume una distribuzione normale per la variabile oggetto di studio e si adotta un valore pari a [*media aritmetica + 2 deviazione standard*] dei risultati generati dal gruppo di pazienti sani. Questo approccio rigido consente di ottenere una specificità molto alta pari al 97.5% (Barajas-Rojas, 1993), ma trascura il valore della sensibilità.

Un altro metodo minimizza la distanza tra l'angolo superiore sinistro, quindi il punto (0,1), e un punto sulla curva ROC (Figura 11). Tale valore rappresenta il miglior compromesso fra sensibilità e specificità e in termini di distanza è pari a

$$d = \sqrt{(1 - Se)^2 + (1 - Sp)^2}.$$

Il punto di cut-off ottimale sarà quello con il valore di d minore in base alla combinazione di Sp e Se (Perkins e Schisterman, 2006).

L'approccio di probabilità di concordanza (Liu, 2012), invece, sceglie il punto di cut-off che massimizza il prodotto delle due corrette probabilità di classificazione, cioè che massimizza la funzione:

$$CP(\tau) = Se(\tau) Sp(\tau),$$

dove $Se(\tau)$ e $Sp(\tau)$ sono rispettivamente la sensibilità e la specificità del test in un punto τ , ossia

$$Se(\tau) = Pr\{Y > \tau\} = 1 - F_Y(\tau),$$

$$Sp(\tau) = Pr\{X \leq \tau\} = F_X(\tau),$$

con $F_X(\cdot)$ che indica la funzione di ripartizione dei pazienti sani X e $F_Y(\cdot)$ la funzione di ripartizione dei pazienti malati Y .

Geometricamente, questo criterio identifica il punto sulla curva ROC che è il vertice del rettangolo con un'area più ampia che si trova sotto la curva (Figura 9).

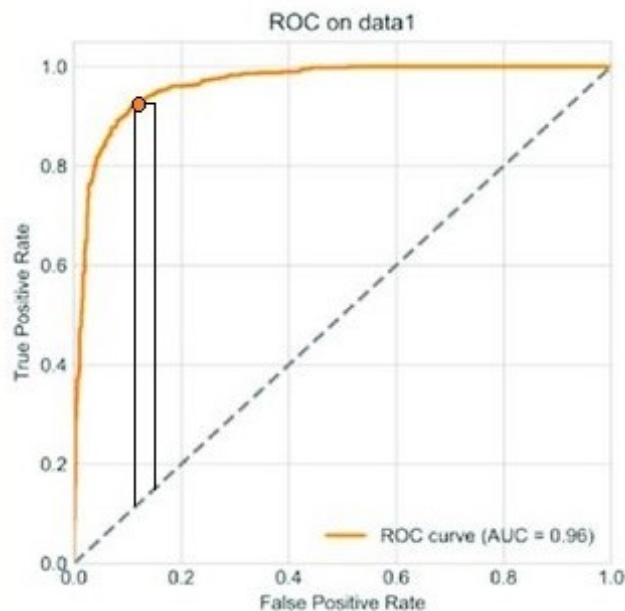


Figura 9: Rappresentazione grafica dell'approccio di probabilità di concordanza

Si può anche rappresentare un grafico con i valori dell'accuratezza, che si ricorda rappresenti la frazione dei casi classificati correttamente, prendendo il valore al quale corrisponde la massima accuratezza (Figura 10).

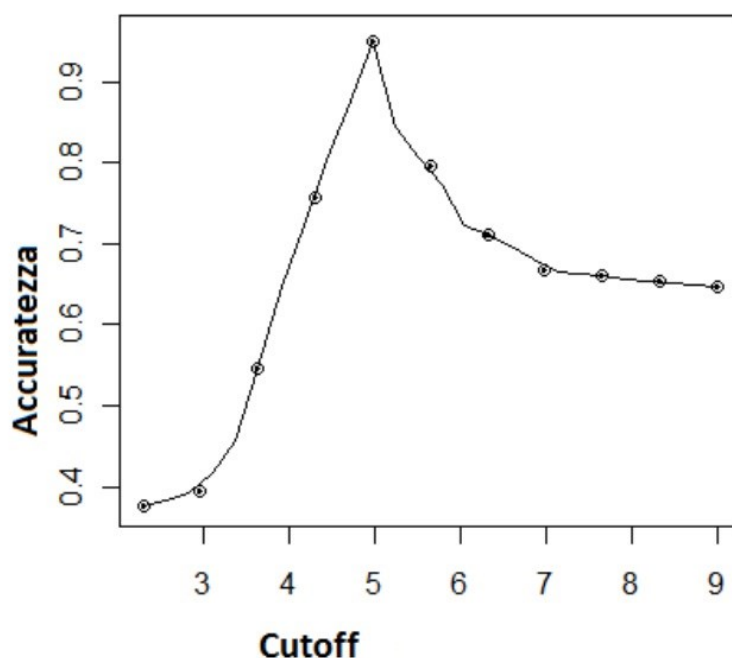


Figura 10: Esempio di scelta del cut-off attraverso l'accuratezza del test. Si vede che il valore massimo di accuratezza si ottiene in corrispondenza del valore di cut-off pari a 5. Tale valore rappresenta il cut-off ottimale.

Degli approcci con l'indice di Youden e sul punto di simmetria se ne parlerà nei prossimi paragrafi.

2.1. *L'indice J di Youden*

Il calcolo del valore del cut-off ottimale tramite la massimizzazione dell'area sottesa alla curva ROC si ritiene che non sfrutti pienamente le caratteristiche adatte a stabilire la soglia migliore. L'attenzione viene posta quindi sulla necessità di trovare quella funzione obiettivo che sia in grado di dare un valore di cut-off il più accurato possibile. Perciò è stato introdotto nella letteratura medica da Youden nel 1950, un indice J (Perkins e Schisterman, 2006), che rappresenta la massimizzazione di una funzione della sensibilità e della specificità rispetto ad un

valore di soglia τ , cioè $J = \max J(\tau)$, che di conseguenza è quel punto che identifica più correttamente la distinzione tra sani e malati. In pratica massimizza la distanza verticale tra la bisettrice (chance line) e il generico punto della curva ROC (Figura 11). L'indice di Youden si determina quindi tramite la seguente funzione:

$$J(\tau) = Se(\tau) + Sp(\tau) - 1,$$

dove $Se(\tau)$ e $Sp(\tau)$ sono, rispettivamente, la sensibilità e la specificità del test nel punto τ . L'indice J è quindi la massima capacità di discriminazione del cut-off in due gruppi quando viene dato lo stesso peso a sensibilità e specificità. Assume valori compresi nell'intervallo chiuso $[0,1]$, dove a valore 0 corrisponde un test completamente inefficace, mentre a valore 1 corrisponde un test perfettamente efficace.

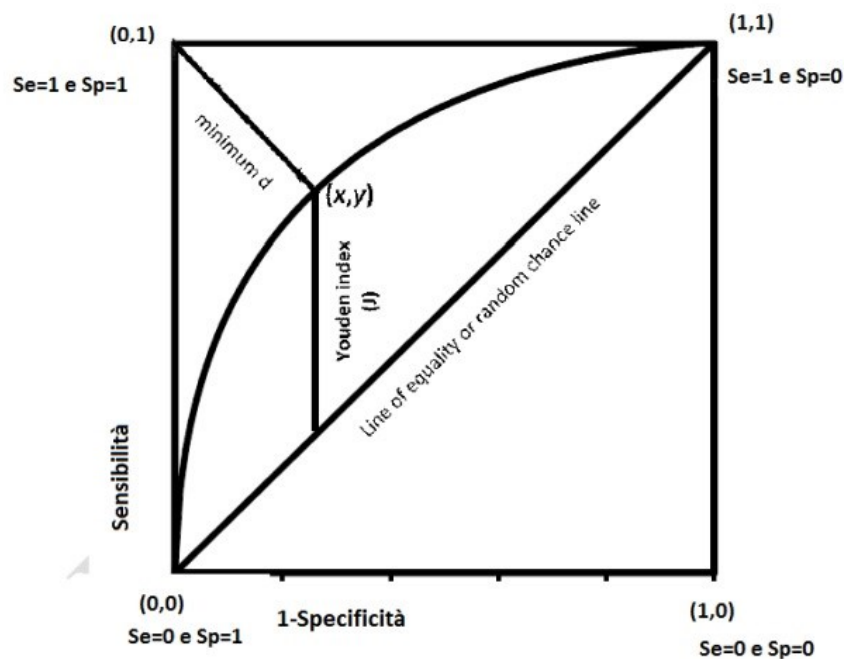


Figura 11: Criterio di Youden e criterio della minima distanza da (0,1).

2.2. Metodo del punto di simmetria ($SP=SE$)

L'approccio del punto di simmetria (Jiménez-Valverde, 2014; López-Ratón, Cadarso-Suárez, Molanes-López e Letón, 2016) controlla simultaneamente la probabilità dei due tipi di classificazione corrette (test negativo come sano e test positivo come malato) e permette di ottenere regioni di confidenza per il cut-off ottimale e il valore della sensibilità associata. Tale metodo consiste nel prendere come cut-off ottimale il punto sulla curva ROC dove $Se(\tau) = Sp(\tau)$, cioè il punto in cui la sensibilità e specificità sono uguali che corrisponde alla probabilità di classificare correttamente qualsiasi soggetto, che sia sano o malato.

A differenza degli altri approcci, quello basato sul punto di simmetria non richiede di risolvere alcun processo di ottimizzazione di una funzione di sensibilità e specificità. Al contrario, come spiegato in López-Ratón *et al.* (2016), da una prospettiva clinica, ci possono essere situazioni in cui è più appropriato controllare contemporaneamente le probabilità di entrambi i tipi di classificazioni corrette piuttosto che massimizzare (o minimizzare) alcune funzioni di esse. Inoltre, dal punto di vista della teoria della decisione bayesiana, l'approccio del punto di simmetria coincide con la regola min-max che mira a minimizzare il massimo della probabilità di errore di classificazione errata. Ciò significa che se τ_s è il cut-off ottimale basato sul punto di simmetria, cioè $Se(\tau_s) = Sp(\tau_s)$, allora

$$\tau_s = \arg \min \{ \max e(\pi, \tau) \},$$

dove $e(\pi, \tau) = (1 - \pi)[1 - Sp(\tau)] + \pi[1 - Se(\tau)]$ e π sono, rispettivamente, la probabilità di errore di classificazione errata e la prevalenza della malattia (quando la prevalenza della malattia $\pi=0.5$, τ_s coincide con il cut-off ottimale prodotto con l'indice di Youden).

Infine, c'è un'interessante generalizzazione dell'approccio del punto di simmetria, che incorpora i costi associati a falsi positivi e falsi negativi. Tale generalizzazione è ancora data da una regola min-max che mira a ridurre al minimo il massimo costo o rischio complessivo previsto. Il cosiddetto approccio generalizzato del punto di simmetria identifica come ottimale il cut-off che soddisfa la seguente equazione:

$$\rho [1 - Sp(\tau)] = 1 - Se(\tau),$$

dove ρ è il costo relativo di una classificazione di falsi positivi rispetto a una classificazione di falsi negativi. Chiaramente, quando $\rho=1$ l'approccio del punto di simmetria generalizzato produce il cut-off ottimale basato sul punto di simmetria standard.

Ci sono diversi lavori che trattano metodi inferenziali per il problema dell'identificazione di una soglia ottimale per i test diagnostici. Tuttavia, ci sono poche proposte nella letteratura statistica per l'inferenza sul punto di simmetria e sul corrispondente cut-off. Per quanto si sa, solo López-Ratón *et al.* (2016) trattano l'argomento e propongono due metodi per la costruzione di distinti intervalli di confidenza per τ_s , la sensibilità $Se_s = Se(\tau_s)$ e specificità $Sp_s = Sp(\tau_s)$ associate. I due metodi proposti sono uno parametrico basato sulla quantità pivotale generalizzata (GPQ) nell'ipotesi di normalità (Weerahandi, 1993, 1995; Lai *et al.*, 2012) che presuppone distribuzione normale per X e Y , e uno non parametrico basato sul valore di verosimiglianza empirico liscio (EL) (Thomas e Grunkemeier, 1975). I metodi di López-Ratón *et al.* (2016) producono intervalli di confidenza distinti (cioè marginali) e sono anche computazionalmente complessi poiché si basano rispettivamente sull'uso di algoritmi Monte Carlo e sulle procedure di ricampionamento.

2.2.1 Inferenza parametrica (GPQ)

Gli intervalli di confidenza generalizzati si riferiscono a una metodologia parametrica basata sull'ipotesi di normalità, introdotta per la prima volta da Weerahandi (1993, 1995) e recentemente applicata nell'ambito della diagnostica per lo studio all'indice Youden di Lai *et al.* (2012) e Zhou e Qin (2013), e al punto di simmetria di López-Ratón *et al.* (2016).

Si ricorda che se si assume che il risultato del test diagnostico nelle due popolazioni sane X e malate Y abbia una distribuzione normale con media μ_x e μ_y , deviazione standard σ_x e σ_y , e ne consegue che la curva ROC è espressa come:

$$ROC(p) = \Phi(a + b\Phi^{-1}(p)), \quad (1)$$

dove $a = \frac{\mu_y - \mu_x}{\sigma_y}$, $b = \frac{\sigma_x}{\sigma_y}$, $p = 1 - Sp(\tau_s)$. Pertanto, nell'ipotesi di normalità, utilizzando se necessario una trasformazione monotona di tipo Box-Cox per soddisfare tale ipotesi, dall'equazione (1) segue che il cut-off ottimale τ_s basato sul punto di simmetria può essere stimato dalla seguente equazione:

$$\Phi(a + b\Phi^{-1}(p)) = 1 - \rho p \Leftrightarrow \Phi\left(\frac{\Phi^{-1}(1 - \rho p) - a}{b}\right) - p = 0, \quad (2)$$

sostituendo a e b con $\tilde{a} = \frac{m_y - m_x}{s_y}$ e $\tilde{b} = \frac{s_x}{s_y}$, rispettivamente, dove m_x , m_y e s_x e s_y sono la media campionaria e la deviazione standard campionaria delle due popolazioni sani e malati. Una volta ottenuta dall'equazione (2) la stima di \tilde{p} , si hanno le stime parametriche puntuali di $\tilde{\tau}_s$, che è il cut-off ottimale, della specificità $\tilde{Sp}(\tau_s)$ e sensibilità $\tilde{Se}(\tau_s)$ ottenute dalle seguenti funzioni:

$$\begin{aligned}\widetilde{\tau}_s &= m_0 + s_0 \Phi^{-1}(1 - \tilde{p}), \\ \widetilde{Sp}(\tau_s) &= (1 - \tilde{p}) \text{ e } \tilde{\theta}(\tau_s) = 1 - \rho\tilde{p},\end{aligned}$$

rispettivamente, dove ρ è il costo relativo di una classificazione di falsi positivi rispetto a una classificazione di falsi negativi.

Per calcolare gli intervalli di confidenza basati su GPQ di τ_s , $Se(\tau_s)$ e $Sp(\tau_s)$, López-Ratón *et al.* (2016) seguono lo stesso ragionamento di Lai *et al.* (2012), basato sull'uso delle quantità pivotali, ovvero sostituendo a e b nell'equazione (2) con le loro corrispondenti quantità pivotali. Ad esempio, se Q_{τ_s} è la quantità pivotale per la stima di τ_s e $Z_{\tau_s, \alpha}$ è l' α -esimo quantile della distribuzione di Q_{τ_s} , il corrispondente intervallo di confidenza per Q_{τ_s} è dato da $Q_{\tau_s} \pm Z_{\tau_s, \alpha/2} \sigma_{\tau_s}$.

2.2.2 Inferenza non parametrica (EL)

Il metodo della verosimiglianza empirica è stato introdotto per la prima volta da Thomas e Grunkemeier (1975) che ha proposto la costruzione di intervalli di confidenza basati su EL (*Empirical likelihood method*) per lo stimatore di Kaplan-Meier. Al giorno d'oggi, questa metodologia è molto utilizzata in diversi campi grazie alle buone proprietà degli intervalli di confidenza e regioni basati su EL. Inoltre, ha il vantaggio di essere facilmente implementabile e non richiede alcuna particolare ipotesi parametrica. Nella letteratura recente, Molanes-López e Letón (2011) hanno proposto un approccio EL basato su bootstrap per fare inferenza sull'indice Youden e sul suo cut-off associato stimando intervalli di confidenza per il punto di simmetria generalizzata e la specificità e sensibilità corrispondente.

Consideriamo un campione casuale da X , (x_1, \dots, x_m) , cioè i risultati del test di m pazienti sani, e un campione casuale (y_1, \dots, y_n) da Y , cioè i risultati del test di n

pazienti malati. Si presume che il vero stato di malattia di ciascun paziente sia accertato da un test golden standard. Inoltre, sia \hat{F}_x la funzione di distribuzione empirica basata su x_1, \dots, x_m e \hat{F}_y la funzione di distribuzione empirica basata su y_1, \dots, y_n . In particolare, quindi, per un valore fisso t ,

$$\hat{F}_X(t) = \frac{1}{m} \sum_{i=1}^m I(x_i \leq t),$$

dove $I(\cdot)$ indica la funzione indicatrice. Ricordiamo che consideriamo un test diagnostico continuo e che, per una soglia fissa τ , $Se = Se(\tau) = 1 - F_y(\tau)$ e $Sp = Sp(\tau) = F_x(\tau)$ e che F_x^{-1} sia la funzione inversa di F_x . A partire dalla relazione $1 - Se = F_y(F_x^{-1}(Sp))$, cioè, le distribuzioni di X e Y devono essere tali che il Sp -quantile τ di X sia uguale al $(1 - Se)$ -quantile di Y , Adimari e Chiogna (2010) definiscono la log-verosimiglianza empirica come (Owen, 2001):

$$\begin{aligned} \ell(Se, Sp, \tau) = & 2m \left\{ \hat{F}_x(\tau) \log \frac{\hat{F}_x(\tau)}{Sp} + [1 - \hat{F}_x(\tau)] \log \frac{1 - \hat{F}_x(\tau)}{1 - Sp} \right\} \\ & + 2n \left\{ \hat{F}_y(\tau) \log \frac{\hat{F}_y(\tau)}{1 - Se} + [1 - \hat{F}_y(\tau)] \log \frac{1 - \hat{F}_y(\tau)}{Se} \right\}, \end{aligned}$$

con $Se \in (0,1)$, $Sp \in (0,1)$, $\tau \in T$, dove $T = [\max\{x_{(1)}, y_{(1)}\}, \min\{x_{(m)}, y_{(n)}\}]$, e $x_{(i)}$ con $i = 1, \dots, m$, e $y_{(j)}$ con $j = 1, \dots, n$. Quando $\tau \notin T$, allora $\ell(\theta, \eta, \tau) = +\infty$. Più precisamente, per ogni τ_0 , $Se_0 = 1 - F(\tau_0)$ e $Sp_0 = F_x(\tau_0)$ quando $\min\{m, n\} \rightarrow +\infty$

$$\ell(Se_0, Sp_0, \tau_0) \xrightarrow{d} \chi_2^2,$$

dove χ_2^2 indica la distribuzione chi-quadrato con due gradi di libertà. Quindi si usa tale risultato per costruire regioni di confidenza non parametriche per una coppia di parametri, per esempio (Se_0, τ_0) , quando Sp_0 è fissato. Perciò

$$\mathcal{R}_\alpha = \{(Se, \tau): \ell(Se, \eta_0, \tau) \leq c_\alpha\},$$

dove $\alpha \in (0,1)$ e c_α è tale che $Pr \{\chi_2^2 > c_\alpha\} = \alpha$ è la regione di confidenza con livello di confidenza $1-\alpha$ per la coppia (sensibilità, cut-off), per un valore fissato della specificità. Tale statistica è però limitata se l'oggetto di inferenza è il parametro τ_s , ovvero il cut-off calcolato con il metodo del punto di simmetria e poiché $Se(\tau_s) = Sp(\tau_s)$, la funzione di log-verosimiglianza $\ell(Se, Sp, \tau)$ può essere scritta come

$$\begin{aligned} \ell_s(Se_s, \tau_s) = 2m \left\{ \widehat{F}_x(\tau_s) \log \frac{\widehat{F}_x(\tau_s)}{Se_s} + [1 - \widehat{F}_x(Se_s)] \log \frac{1 - \widehat{F}_x(\tau_s)}{1 - Se_s} \right\} \\ 2n \left\{ \widehat{F}_Y(\tau_s) \log \frac{\widehat{F}_Y(\tau_s)}{1 - Se_s} + [1 - \widehat{F}_Y(\tau_s)] \log \frac{1 - \widehat{F}_Y(\tau_s)}{Se_s} \right\} \end{aligned} \quad (3)$$

per $Se_s \in (0, 1)$ e $\tau_s \in T$. La funzione si adatta al nuovo vincolo $Se_s = Sp_s$ e il risultato asintotico risulta essere:

$$\ell_s(Se_s, \tau_s) \xrightarrow{d} \chi_2^2$$

quando $\min\{m, n\} \rightarrow +\infty$, sotto i veri valori dei parametri. Di conseguenza, l'insieme

$$\mathcal{R}_{S\alpha} = \{(Se_s, \tau_s): \ell(Se_s, \tau_s) \leq c_\alpha\},$$

è una regione di confidenza con livello di significatività $1 - \alpha$ per la coppia di parametri τ_s che è il cut-off ottimale basato sul punto di simmetria e il valore associato di sensibilità Se_s .

Da un punto di vista pratico, una regione congiunta fornisce informazioni più precise sulla coppia di parametri di interesse rispetto agli intervalli di confidenza marginali. Le regioni ottenute da $\ell_s(Se_s, \tau_s)$ sono molto semplici da calcolare, perché coinvolgono essenzialmente solo le funzioni di distribuzione empirica \hat{F}_x e \hat{F}_y ma poiché non sono funzioni lisce, le regioni di confidenza hanno estremi approssimativi.

Se invece la specificità Sp non è nota a priori, Lopez-Ratón *et al.* (2016) propongono di stimare questo parametro prima per mezzo del metodo di Kernel e quindi ottenere una stima puntuale non parametrica di τ minimizzando la log-verosimiglianza e sostituendo la specificità con la sua stima.

Inoltre, Lopez-Ratón *et al.*, introducono un approccio EL basato sul bootstrap che riproduce questa procedura per ogni coppia di valori presi dai due campioni iniziali e costruisce intervalli di confidenza basati su tutte le stime calcolate in precedenza per tali bootstrap.

Infine, altre misure che possono essere utilizzate per descrivere le prestazioni di un test diagnostico sono i cosiddetti rapporti di verosimiglianza. Il rapporto di probabilità positiva a un dato cut-off τ è definito come il rapporto tra la probabilità di classificare correttamente un paziente malato e la probabilità di classificare erroneamente un paziente non malato. Il rapporto di verosimiglianza negativa, invece, è il rapporto tra la probabilità di classificare erroneamente un paziente malato e la probabilità di classificare correttamente un paziente non malato. Quindi, i rapporti di verosimiglianza sono funzioni della sensibilità e della specificità del test. Più precisamente, per i rapporti di probabilità positivi, $\lambda^+(\tau)$, e negativi, $\lambda^-(\tau)$, abbiamo

$$\lambda^+(\tau) = \frac{Se(\tau)}{1-Sp(\tau)} \text{ e } \lambda^-(\tau) = \frac{1-Se(\tau)}{Sp(\tau)}$$

rispettivamente. Quindi, possiamo esprimere la sensibilità Se e la specificità Sp come funzioni di λ^+ e λ^-

$$Se = \frac{\lambda^+(1 - \lambda^-)}{\lambda^+ - \lambda^-}, Sp = \frac{\lambda^+ - 1}{\lambda^+ - \lambda^-}.$$

Poiché se $\tau = \tau_s$ abbiamo $\lambda^+_{s} = 1/\lambda^-_{s}$, con $\lambda^+_{s} = \lambda^+(\tau_s)$ e $\lambda^-_{s} = \lambda^-(\tau_s)$, la funzione (3) può essere formulata, ad esempio, in termini di λ^+_{s} e τ_s sostituendo Se_s con $\lambda^+_{s}/(\lambda^+(\tau_s) + 1)$ e può essere utilizzata per ottenere regioni di confidenza congiunte non parametriche per il cut-off ottimale basato sul punto di simmetria e il rapporto di verosimiglianza positivo associato.

2.2.3. Studio di una simulazione

Si considera un modello parametrico normale per generare le risposte ai test di pazienti sani e malati, in particolare, si considera la seguente distribuzione

(a) $X \sim N(\mu_x = 0, \sigma_x^2 = 1)$ e $Y \sim N(\mu_y, \sigma_y^2 = 0.25)$ ((Biometrical Journal 2020, G.Adimari).

Per eseguire esperimenti di simulazione sull'accuratezza delle regioni di confidenza usando $\ell_s(Sp_s, \tau_s)$, vengono generati campioni casuali per i valori X e Y scegliendo i valori per μ_y , quelli scelti per la specificità Sp_s sono invece, 0.7, 0.8 e 0.9. Per quanto riguarda le dimensioni del campione, per ogni valore di riferimento Sp_s , si sono scelte diverse numerosità. Ogni esperimento di simulazione si basa su 10.000 repliche.

τ_s	μ_Y	Sp	m	n	0.90	0.95	0.99
0.524	0.786	0.7	20	20	0.880	0.950	0.990
			50	20	0.887	0.947	0.988
			20	50	0.884	0.942	0.988
			50	50	0.896	0.950	0.989
			100	100	0.900	0.948	0.990
0.842	1.263	0.8	25	25	0.887	0.950	0.986
			50	25	0.895	0.948	0.989
			25	50	0.893	0.944	0.988
			50	50	0.898	0.948	0.990
			100	100	0.899	0.947	0.988
1.282	1.923	0.9	50	50	0.877	0.945	0.983
			50	75	0.887	0.941	0.987
			75	50	0.884	0.938	0.985
			100	100	0.894	0.944	0.988

Tabella 2: Quantili stimati delle regioni di confidenza per (Sp_s, τ_s) quando $X \sim N(\mu_x = 0, \sigma_x^2 = 1)$ e $Y \sim N(\mu_y, \sigma_y^2 = 0.25)$, con livello di confidenza 0.9, 0.95, 0.99.

Per i tre livelli di confidenza $1 - \alpha$, cioè 0.90, 0.95 e 0.99, la tabella 2, fornisce le probabilità stimate delle regioni di confidenza ottenute utilizzando il metodo della massima verosimiglianza. I dati utilizzati sono stati simulati dalla distribuzione normale e i risultati mostrano che le regioni di confidenza basate su $\ell_s (Sp_s, \tau_s)$ sono sufficientemente accurate nei casi considerati, anche per le dimensioni del campione più piccole. Come previsto, la precisione aumenta con le dimensioni del campione, chiaramente, più i valori di Sp_s sono vicini a 1, maggiori sono le dimensioni del campione richieste.

CAPITOLO 3

Applicazione dei metodi per la ricerca del cut-off

In questo capitolo si ricerca il cut-off ottimale con i metodi menzionati nei capitoli precedenti in due dataset; il primo sulla leishmaniosi canina e il secondo sui linfomi. In particolare, si stima la soglia ottimale inizialmente con il metodo del punto di simmetria confrontando l'approccio parametrico con quello non parametrico, e successivamente verrà utilizzato il criterio dell'indice di Youden, il criterio della minima distanza dal punto ottimale (0,1) e il metodo basato sul grafico dell'accuratezza. A tale proposito tramite il software R, si utilizza la funzione "*gsym.point*" della libreria *GsymPoint* per il metodo del punto di simmetria. Quest'ultimo è un pacchetto che consente di ottenere stime puntuali e intervallari del cut-off ottimale e i relativi indici di sensibilità e specificità in quel punto, con il metodo del punto di simmetria utilizzando la quantità pivotale generalizzata (GPQ) basata sull'ipotesi di normalità e l'approccio non parametrico della verosimiglianza empirica (EL). Per il criterio dell'indice di Youden e della distanza minima dal punto (0,1), invece, si utilizza la funzione "*coords*" della libreria *pROC*, che restituisce le coordinate della curva ROC nel punto di cut-off con il metodo selezionato. Mentre per il grafico dell'accuratezza si utilizzano le funzioni "*prediction*" e "*performance*" della libreria *ROCR* che servono per valutare la classificazione delle variabili creando una previsione dei dati.

3.1. Esempio: Leishmaniosi canina

Si considera uno studio su 30 soggetti relativo alla leishmaniosi canina che è una malattia che provoca un problema renale nei cani. I soggetti vengono classificati dalla variabile "LG" (leishvet grade) a seconda dello stato clinico malato ($LG=1$) o sano ($LG=2$). L'obiettivo è trovare il cut-off ottimale che discrimina meglio i soggetti nei due gruppi utilizzando il metodo del punto di simmetria, il criterio dell'indice di Youden, il criterio della minima distanza da (0,1) e il grafico dell'accuratezza.

Le variabili su cui si basa il test diagnostico sono le proteine *UPC0* e la concentrazione di amilasi nelle urine *UrinaryAmilase0*. Da un'analisi preliminare si nota che nei gruppi dei malati si hanno valori meno elevati per la variabile *UPC0*, così come per la variabile *UrinaryAmilase0* (Figura 12).

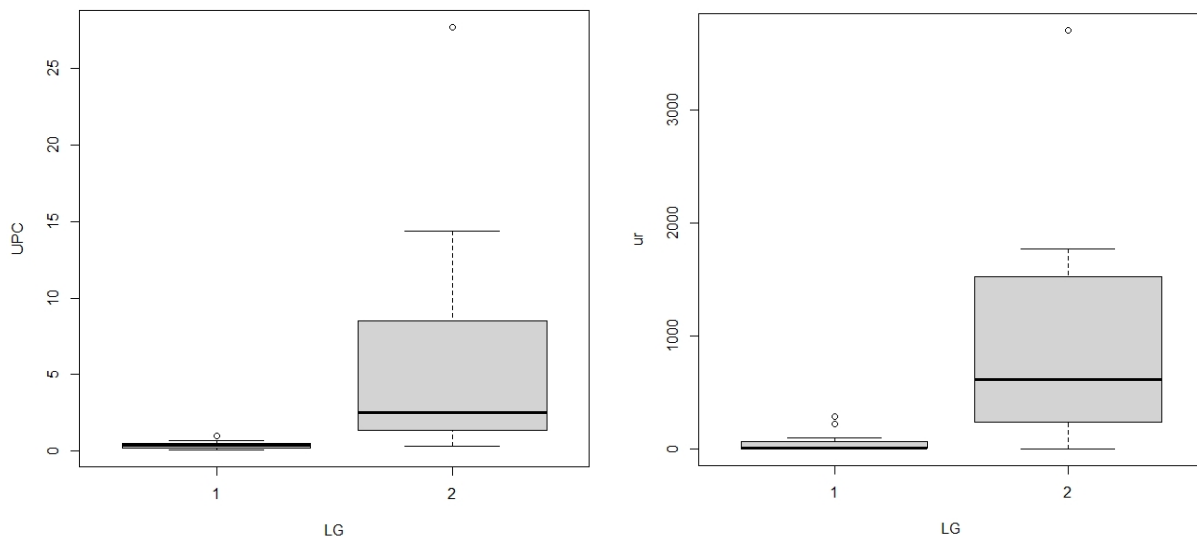


Figura 12: Rappresentazione grafica della distribuzione delle variabili nei due gruppi UPC0 (a sx) e UrinaryAmilase0 (a dx)

Studiando la distribuzione delle variabili *UPC0* e *UrinaryAmilase0* si rifiuta per il test di Shapiro-Wilk l'ipotesi di normalità poiché si hanno $p\text{-value} < 0.01$ (Tabella 3). Tuttavia, applicando una trasformazione logaritmica alle variabili, l'ipotesi di

normalità viene accettata con il test di Shapiro-Wilk per la variabile UPC0, in quanto il pvalue è >0.01 , ma non per la variabile UrinaryAmilase0 (pvalue=0.003), quindi si applica una trasformazione Box-Cox con $\hat{\lambda} = 0.1719$, ottenendo così normalità per entrambe le variabili (Tabella 3).

	p-value Test Shapiro-Wilk		p-value Test Shapiro-Wilk con le variabili trasformate	
	sani	<i>malati</i>	<i>sani</i>	<i>malati</i>
UPC0	0.0015	0.2063	0.5813	0.4882
UrinaryAmilase0	0.013	<0.001	0.0736	0.1404

Tabella 3: Valori del p-value dati originali e dati trasformati con trasformazione logaritmica per UPC0 e trasformazione Box-Cox per UrinaryAmilase0.

Si applica quindi il metodo parametrico del punto di simmetria (Tabella 4), ottenendo così i valori del cut-off ottimale e della sensibilità relativa alla soglia ottimale. Quindi la specificità in tale punto coincide con la sensibilità stimata e ai relativi intervalli di confidenza. In corrispondenza della soglia ottimale del test diagnostico basato sulla variabile trasformata *UPC0* si ha un valore pari a $\hat{\tau}_s = 0.695$, con intervallo di confidenza (0.45, 1.07) e si ha una specificità e sensibilità del 86%. Per il test diagnostico basato sulla variabile *Urinaryamilase0*, invece, si sceglie un cut-off pari a $\hat{\tau}_s = 73.27$, I.C. (28.65, 175.84), con sensibilità del 82%.

	$\hat{\tau}_s$	I.C. $\hat{\tau}_s$	<i>Se</i> ($\hat{\tau}_s$)	I.C. <i>Se</i> ($\hat{\tau}_s$)
Log(UPC0)	0.695	0.45 - 1.07	0.86	0.72-0.95
UrinaryAmilase0	73.27	28.65-175.84	0.82	0.68-0.91

Tabella 4: Stima del cut-off e sensibilità con il metodo parametrico GPQ.

Se si vuole utilizzare l'approccio non parametrico della verosimiglianza empirica (EL) per la ricerca del cut-off ottimale con il metodo del punto di simmetria, si applica la stessa funzione "*gsym.point*" specificando il metodo "*EL*", ottenendo i

valori riportati nella Tabella 5. Nella Figura 13 e 14 sono rappresentati graficamente i cut-off ottimali scelti con il metodo parametrico e non parametrico del punto di simmetria.

	$\hat{\tau}_s$	I.C. $\hat{\tau}_s$	$Se(\hat{\tau}_s)$	I.C. $Se(\hat{\tau}_s)$
UPCO	0.70	0.42-0.99	0.86	0.65-0.93
UrinaryAmilase0	100	21.28-250	0.82	0.67-0.93

Tabella 5: Stima del cut-off e sensibilità con il metodo non parametrico EL.

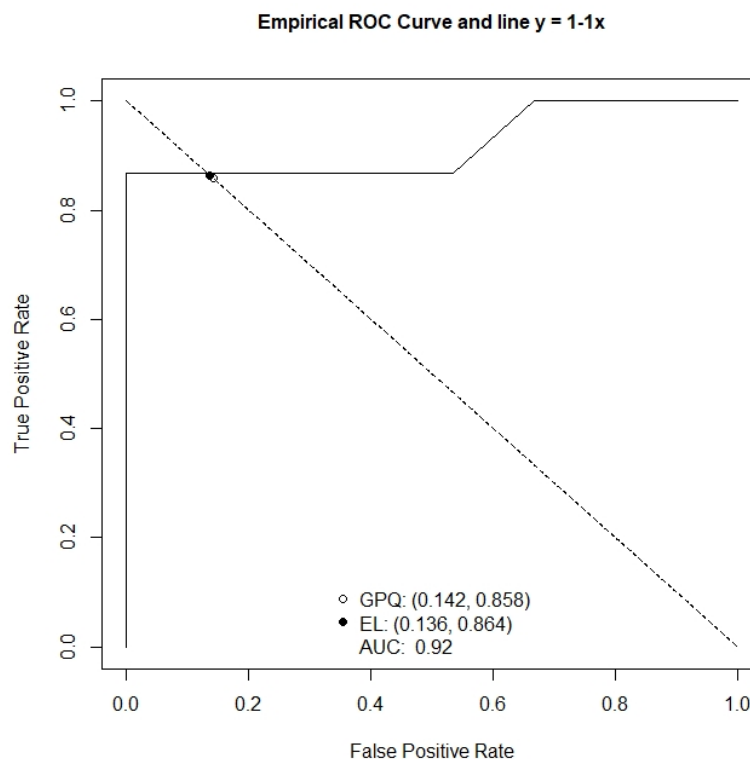


Figura 13: Rappresentazione grafico curva ROC con cut-off ottimale scelto con il metodo parametrico e non parametrico del punto di simmetria per la variabile UPCO.

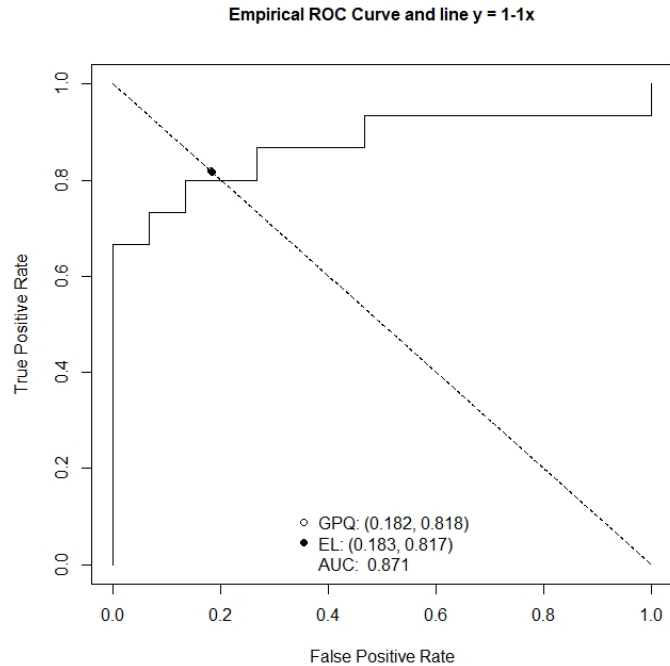


Figura 14: Rappresentazione grafico curva ROC con cut-off ottimale scelto con il metodo parametrico e non parametrico del punto di simmetria per le variabili *UrinaryAmilase0*.

Come si può notare dalle Tabelle 4 e 5 e dalle Figure 13 e 14, i valori risultano essere molto simili sia per il cut-off ottimale che per la sensibilità (=specificità) utilizzando l'approccio parametrico GPQ (Tabella 4) e non parametrico EL (Tabella 5). Tuttavia, se le variabili o le loro trasformazioni hanno distribuzione normale, si preferisce il metodo parametrico (GPQ), altrimenti si utilizza il metodo non parametrico (EL). La funzione "*gsym.point*" permette di scegliere l'approccio che meglio può essere applicato ai dati scegliendo come metodo "auto". In questo caso, poiché le variabili *UPC0* e *UrinaryAmilase0* dopo essere state trasformate hanno distribuzione normale, si utilizza il metodo parametrico (GPQ).

Utilizzando come metodi alternativi il criterio dell'indice di Youden e della minima distanza dal punto (0,1) si specifica nella funzione "*coords*" il metodo non parametrico "*youden*" e "*closest.topleft*". Entrambi gli approcci confermano la scelta della soglia ottimale $\hat{t} = 1.1$ per la variabile *UPC0* e $\hat{t} = 150.8$ per la variabile *UrinaryAmilase0*. In corrispondenza di tali valori si ottiene una specificità del 100%

e una sensibilità del 86% per la variabile *UPC0*, una specificità del 87% e sensibilità del 80% per la variabile *UrinaryAmilase0* (Tabella 6).

	\hat{t}	Sp (\hat{t})	Se (\hat{t})
UPC0	1.1	1	0.86
UrinaryAmilase0	150.8	0.87	0.81

Tabella 6: Stima del cut-off, specificità e sensibilità con il criterio di Youden e distanza minima da (0,1).

Dalla Figura 15, invece, possiamo stimare dal grafico dell'accuratezza il cut-off ottimale per la variabile *UPC0* che approssimativamente risulta essere compreso tra 0 e 5 e per la variabile *UrinaryAmilase0* tra 100 e 200, in corrispondenza del valore di massima accuratezza.

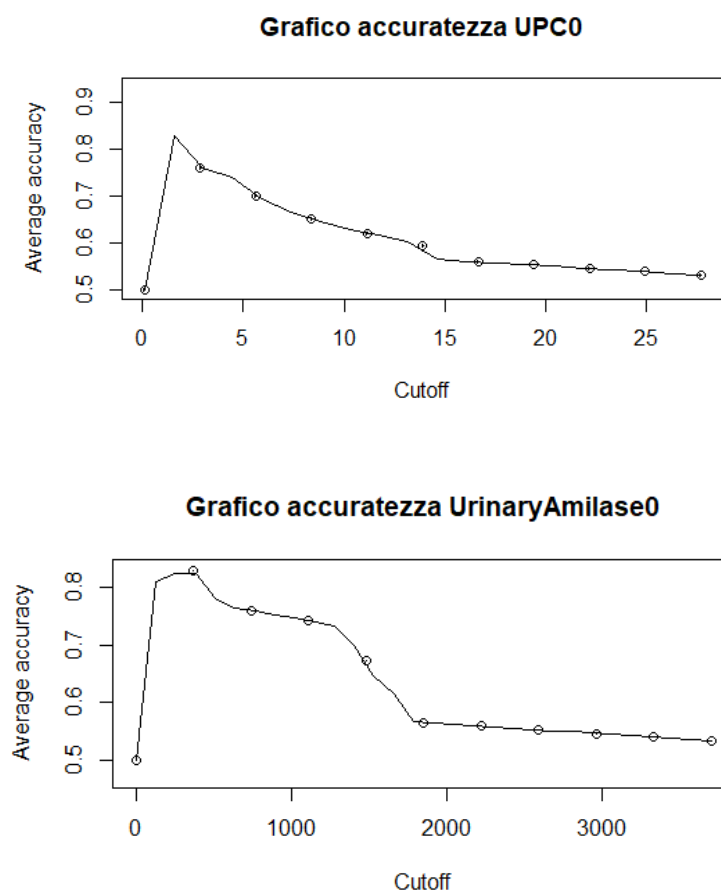


Figura 15: Grafici accuratezza delle variabili *UPC0* (in alto) e *UrinaryAmilase0* (in basso)

In conclusione, si può dire che approssimativamente i metodi utilizzati per la ricerca del cut-off ottimale, individuano valori piuttosto simili per la variabile UPC0; in particolare con il metodo del punto di simmetria con i due approcci, parametrico ($\hat{\tau}_s = 0.695$) e non parametrico ($\hat{\tau}_s = 0.70$), si hanno valori che approssimativamente coincidono. Per la variabile UrinaryAmilase0, invece, si hanno valori diversi; in particolare con l'approccio parametrico (GPQ) del punto di simmetria si ha un valore del cut-off minore ($\hat{\tau}_s = 73.27$) rispetto all'approccio non parametrico ($\hat{\tau}_s = 100$) e del criterio di Youden, che ha lo stesso valore del criterio della distanza minima dal punto (0,1) ($\hat{\tau} = 150.8$). Si preferisce comunque utilizzare l'approccio parametrico poiché entrambe le variabili trasformate si distribuiscono normalmente.

3.2. Esempio: Linfomi

Si considera un campione di 55 soggetti divisi in due gruppi: 18 soggetti malati e 37 soggetti sani. Si eseguono analisi esplorative sui dati osservando che i microRNA che discriminano meglio i pazienti nei due gruppi risultano essere rappresentati dalle variabili *miR20a* e *miR92a* (Figura 16) (Ventura e Racugno, 2017).

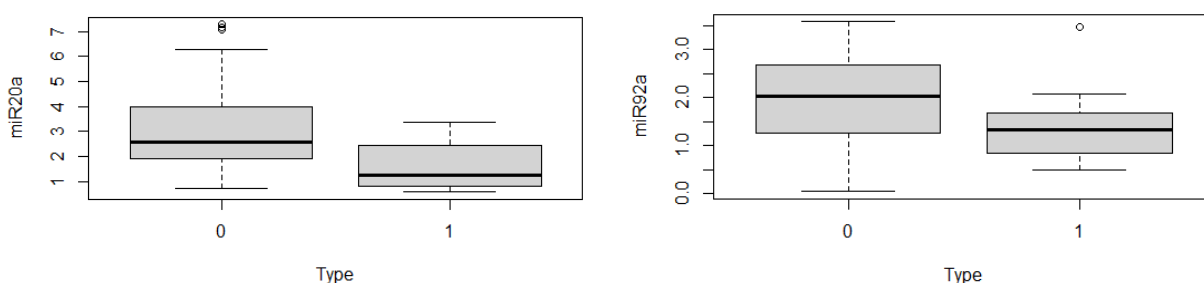


Figura 16: Rappresentazione grafica distribuzione della variabile miR20a e miR92a nei due gruppi.

Per il calcolo del cut-off ottimale con il metodo del punto di simmetria si procede come nell'esempio precedente utilizzando la funzione *gsym.point* della libreria *GsymPoint* su R dopo aver studiato la normalità delle distribuzioni delle variabili con il test di Shapiro Wilk. Poiché i p-value sono <0.01 , le variabili non hanno distribuzione normale, perciò si esegue una trasformazione logaritmica per la variabile miR20a e una trasformazione Box-Cox con $\hat{\lambda}=0.74$ per la variabile miR92a per ottenere così due distribuzioni normali (Tabella 6).

	p-value Test Shapiro-Wilk		p-value Test Shapiro-Wilk con variabile trasformata	
	<i>sani</i>	<i>malati</i>	<i>sani</i>	<i>malati</i>
miR20a	0.03	0.0023	0.765	0.1564
miR92a	0.025	0.5407	0.1408	0.2465

Tabella 7: Valori del p-value dati originari e dati trasformati con trasformazione logaritmica per miR20a e trasformazione Box-Cox per miR92a.

Dalla Tabella 7 si può notare che le variabili trasformate soddisfano l'ipotesi di normalità; quindi, si utilizza l'approccio parametrico (GPQ) del metodo del punto di simmetria per scegliere il cut-off ottimale e ottenere il valore della sensibilità in quel punto (Tabella 8).

	$\hat{\tau}_s$	I.C. $\hat{\tau}_s$	$Se(\hat{\tau}_s)$	I.C. $Se(\hat{\tau}_s)$
miR20a	1.93	1.59-2.3	0.73	0.651-0.82
miR92a	1.58	1.34-1.84	0.63	0.55-0.74

Tabella 8: Stima del cut-off e sensibilità con il metodo parametrico GPQ.

Quindi la soglia ottimale per la variabile *miR20a* $\hat{\tau}_s=1.93$, con I.C.(1.59,2.3) e sensibilità (=specificità) del 73%. Mentre per la variabile *miR92a* il cut-off ottimale è pari a $\hat{\tau}_s=1.58$, con I.C. (1.34,1.84) con sensibilità del 63%.

Utilizzando il criterio dell'indice di Youden e il criterio della massima distanza, invece, si sceglie con entrambi i metodi un cut-off pari a $\hat{\tau} = 1.29$ per miR20a con specificità del 92% e una sensibilità del 61% in corrispondenza di tale valore. Mentre per miR92a si sceglie un cut-off ottimale con il criterio di Youden pari a $\hat{\tau} = 1.88$ con specificità del 57% e una sensibilità del 88%, e con il criterio della minima distanza dal punto (0,1), si sceglie $\hat{\tau} = 1.78$ con specificità del 59% e sensibilità del 83% (Tabella 9).

Criterio di Youden	$\hat{\tau}$	Sp ($\hat{\tau}$)	Se ($\hat{\tau}$)
miR20a	1.29	0.92	0.61
miR92a	1.88	0.88	0.57
Criterio della minima distanza da (0,1)			
miR20a	1.29	0.92	0.61
miR92a	1.78	0.59	0.83

Tabella 9: Stima del cut-off , specificità e sensibilità con il criterio di Youden e distanza minima da (0,1).

Dal grafico dell'accuratezza, invece, si sceglie il cut-off ottimale in corrispondenza del massimo valore. Quindi per la variabile miR20a il cut-off ottimale è compreso tra 1 e 2, mentre per la variabile mir92a è compreso tra 1 e 1.5 (Figura 16).

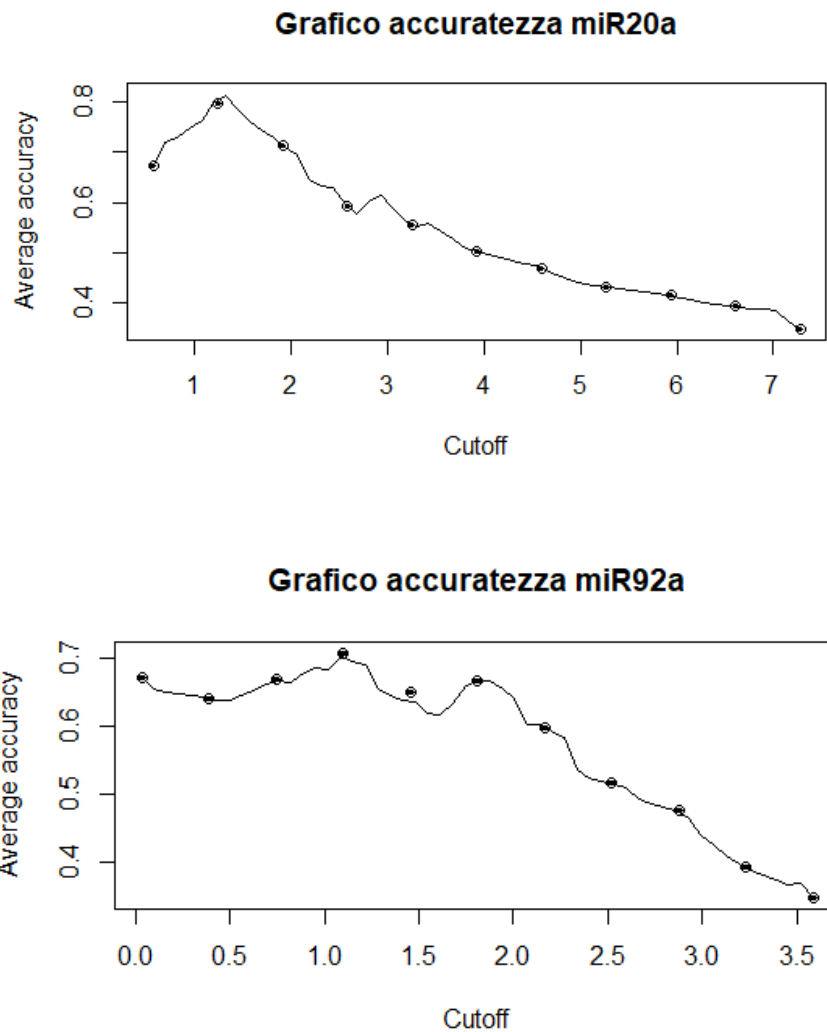


Figura 16: Grafici accuratezza delle variabili miR20a (in alto) e miR92a (in basso)

In conclusione, si può dire che con il metodo parametrico si ha un valore più elevato del cut-off per la variabile miR20a ($\hat{\tau}_s=1.93$) rispetto al criterio di Youden ($\hat{\tau} = 1.29$) che coincide con il criterio della minima distanza dal punto (0,1). Per la variabile miR92a, invece si ha un valore più elevato con il criterio di Youden ($\hat{\tau} = 1.88$), rispetto al criterio della minima distanza dal punto (0,1) ($\hat{\tau} = 1.78$) e all'approccio parametrico del punto di simmetria ($\hat{\tau}_s=1.93$). Come per l'esempio precedente, si preferisce utilizzare il metodo del punto di simmetria con l'approccio parametrico poiché entrambe le variabili trasformate si distribuiscono normalmente.

CONCLUSIONI

In conclusione, si può dire che per valutare l'accuratezza di un test con valori continui si utilizza la curva ROC che rappresenta i vari livelli discriminanti di un esame diagnostico e permette di identificare il valore della soglia ottimale, il cosiddetto "best cut-off". Ci sono diversi metodi per calcolare il cut-off ottimale di un test e la relativa sensibilità e specificità in tale punto. In questa tesi vengono trattati diversi metodi come: il criterio della minima distanza, che minimizza la distanza tra l'angolo superiore sinistro, il punto (0,1), e un punto sulla curva ROC; il metodo di probabilità di concordanza che massimizza il prodotto delle due corrette probabilità di classificazione e il metodo grafico che rappresenta i valori dell'accuratezza scegliendo il punto al quale corrisponde il massimo valore. Inoltre, ci si sofferma sul metodo dell'indice di Youden, che rappresenta la massimizzazione di una funzione della sensibilità e specificità rispetto ad un valore soglia τ che di conseguenza corrisponde al cut-off ottimale e al metodo del punto di simmetria. Per quest'ultimo esiste un approccio parametrico basato sulla quantità pivotale generalizzata (GPQ) e un approccio non parametrico basato sul valore della verosimiglianza empirica (EL). Nell'ultimo capitolo vengono poi applicati alcuni metodi parametrici e non parametrici a due dataset, concludendo che se le variabili, o le loro trasformazioni, si distribuiscono normalmente si preferisce l'approccio parametrico rispetto a quello non parametrico.

Bibliografia

- Adimari, G. & Chiogna, M. (2010) Simple nonparametric confidence regions for the evaluation of continuous-scale diagnostic tests. *The International Journal of Biostatistics*.
- Bamber, D. (1975), *The area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph*, J.
- Bottarelli, E., Parodi, S., (2003), Un approccio per la valutazione della validità dei test diagnostici: le curve R.O.C. (*Receiver Operating Characteristic*).
- Kotz, S., Lumelskii, Y., Pensky, M. (2003), *The Stress-Strength Model and its Generalizations. Theory and Applications*. World Scientific, Singapore.
- Liu, Aiyi Liu, Enrique Schisterman, Yan Zhu. On linear combinations of biomarkers to improve diagnostic accuracy. *Stat Med* 2005.
- Metz, C.E (1978), *Basic principles of ROC analysis, Seminar Nuclear Medicine*, VIII, 283-29.
- Pace, L., Salvan, A. (2001), *Introduzione alla Statistica: Inferenza, verosimiglianza, modelli*. Cedam, Padova.
- Perkins, E. Schisterman *Medicine American journal of epidemiology* 2006
- Swets, J.A. (1998), *Measuring the accuracy of diagnostic systems*.
- Schisterman EF, Perkins N. Confidence intervals for the youden index and corresponding optimal cut point. 2007.
- Yin JY, Lili Tian. Optimal linear combinations of multiple diagnostic biomarkers based on youden index. *Stat Med* 2013.
- S. Weerahandi. *Exact Statistical Methods for Data Analysis*. Springer-Verlag, 1995.