Ratio Mathematica 11
(1996), 35 - 45

LA DIAGNOSTICA DIFFERENZIALE MEDIANTE L'ANALISI DISCRIMINANTE

Enzo Ballone*, Vittorio Colagrande**

SUNTO - Gli Autori hanno applicato tecniche di analisi discriminante ai risultati di alcuni test di laboratorio nella diagnostica differenziale tra epatite virale e ittero colostatico nelle fasi iniziali di queste malattie. I test utilizzati come *caratteri* nella discriminazione sono: transaminasi glutammico-ossalacetico (GOT), transaminasi glutammico-piruvico (GPT) e fosfatasi alcalina (FA). Combinando variamente i risultati dei test, sono state calcolate le diverse *funzioni discriminanti*. La "migliore discriminazione" diagnostica è stata ottenuta calcolando la funzione discriminante sui dati campionari di GOT, GPT e FA del siero considerati congiuntamente.

ABSTRACT - The Authors applied the discriminant analysis calculation to the results of some laboratory tests in the differential diagnosis between viral hepatitis and colostatic icterus in the starting phases of the said diseases. The laboratory tests used as *characters* in the discrimination were: glutamic-oxaloacetis transaminase (GOT), glutamic-pyruvic transaminase (GPT) and alkaline phosphatasis (AP). Many different *discriminant functions* were calculated by combining in different ways the laboratory tests considered. The best diagnostic discrimination was obtained by calculating the discriminant function on the results of GOT, GPT and serum AP determined simultaneously.

* Cattedra di Statistica Medica, Università "G. D'Annunzio" - Chieti.
** Liceo Scientifico "G. Galilei" - Lanciano (CH)

INTRODUZIONE

L'Analisi Discriminante è una metodologia matematico-statistica, introdotta per la prima volta da R.A. Fisher nel 1936 e applicabile principalmente a caratteri quantitativi, che permette una valutazione "globale" e "bilanciata" dei diversi caratteri presi in esame, la separazione di elementi in gruppi distinti e l'attribuzione di una qualsiasi nuova unità ad uno dei gruppi definiti.

La presente nota propone dei modelli per "discriminare" tra due popolazioni in base ai valori di alcune variabili statistiche. Oltre a presentare il problema della discriminazione, viene offerto lo spunto per riflettere, in particolare, sull'incertezza nell'attribuzione di un elemento a gruppi distinti definiti in precedenza. E questa tematica si presenta quotidianamente, ad esempio, a chi lavora nell'ambito della diagnostica medica per la differenziazione tra soggetti "normali" e non, tra quelli affetti da una patologia o da un'altra che presentano gli stessi sintomi.

MATERIALI E METODI

Vengono riportati i risultati dell'analisi discriminante condotta su 40 casi di epatite virale (EV) e 30 casi di ittero colostatico (IC). Nella casistica sono stati inclusi solamente quei pazienti per i quali la diagnosi è stata effettuata con certezza, indipendentemente dalle prove di laboratorio, attraverso la biopsia epatica per l'epatite virale ed i reperti operatori per l'ittero colostatico.

In ciascuno dei pazienti presi in esame, su un unico campione di siero prelevato non oltre 15 gg. dall'insorgenza dell'ittero, sono state eseguite le determinazioni dei tassi serici delle transaminasi glutammico-ossalacetico (GOT) e glutammico-piruvico (GPT) nonché della fosfatasi alcalina (FA). Atteso che questi enzimi aumentano, con valori massimi fino a 500 volte superiore alla norma (specialmente la GPT), prima della comparsa dell'ittero nelle EV e più tardivamente ed in modo moderato nell'IC, si pone il problema della diagnosi differenziale tra EV ed IC. L'analisi discriminante, applicata a questi dati di laboratorio, ha permesso di differenziare tra pazienti affetti da EV e pazienti affetti da IC, minimizzando l'errore di discriminazione. Sono stati determinati, inoltre, i coefficienti di correlazione relativi a tutte le associazioni possibili tra i vari caratteri presi in esame, sia nel gruppo di pazienti affetti da EV che in quello dei pazienti affetti da IC. La *soglia discriminante ottimale* relativa ad ogni singolo carattere registrato è stata stimata, in un primo step, come media aritmetica delle due medie del carattere stesso nei due gruppi di pazienti e successivamente come media ponderata con le varianze. Il calcolo delle *funzioni discriminanti* e dei relativi errori percentuali di "misclassificazione", per le

variabili prese a coppie e considerate tutte, è stato eseguito mediante il software SPSS. Il tutto sulla base di alcune metodologie matematico-statistiche che si illustreranno di seguito.

Prese in esame due popolazioni U_1 e U_2, sia X un carattere quantitativo rilevato su di esse; X verrà assimilato ad una variabile casuale avente funzioni di densità di probabilità $f_1(x)$ su U_1 e $f_2(x)$ su U_2. Si supponga di estrarre due campioni casuali indipendenti, l'uno da U_1 e l'altro da U_2, rispettivamente di n_1 ed n_2 elementi, con medie aritmetiche $\bar{x}_1$ e $\bar{x}_2 0$ (con $\bar{x}_1 < \bar{x}_2 0$) e deviazioni standard s_1 e s_2.

Un *valore discriminante (o soglia)* c della variabile, con $\bar{x}_1 < c < \bar{x}_2 0$, è tale da suddividere lo spazio $A = (-\infty, +\infty)$ dei valori di X in due sottospazi $A_1 = (-\infty, c)$ ed $A_2 = (c, +\infty)$ tali che se una nuova osservazione appartiene a A_1, essa viene attribuita alla popolazione U_1, altrimenti a U_2. Per determinare c si possono ipotizzare due situazioni diverse: a) le probabilità a priori p_1 e p_2 che una generica osservazione provenga da U_1 o da U_2 sono note; b) le probabilità a priori non sono note. In realtà sarebbe necessario considerare anche il *peso* (o *costo*) g_{ij} di errata assegnazione di un individuo della popolazione U_k alla U_h, ma nel seguito si supporrà $g_{kk} = 0$ e $g_{kh} = 1$ per $k \neq h$.

Per la situazione a), l'errore che si commette assegnando a U_1 un'osservazione di U_2 e quello di attribuzione ad U_2 di un'osservazione di U_1 sono dati, rispettivamente, dalle probabilità condizionate:

$$\varepsilon_2(c) = p(X < c|U_2) = \int_{-\infty}^{c} f_2(x)\,dx, \quad \varepsilon_1(c) = p(X > c|U_1) = \int_{c}^{+\infty} f_1(x)\,dx.$$

L'errore complessivo atteso di assegnazione risulta allora

$$\varepsilon(c) = p(U_1)\,p(X > c|U_1) + p(U_2)\,p(X < c|U_2) = p_1\varepsilon_1(c) + p_2\varepsilon_2(c).$$

La soglia c va ricercata tra i valori di X che minimizzano l'errore atteso, ovvero la funzione $\varepsilon(x)$. Derivando questa sia ha: $\varepsilon'(x) = p_2\,f_2(x) - p_1\,f_1(x)$ e si ottiene un minimo prendendo c tale che

$$\begin{cases} p_2\,f_2(x) - p_1\,f_1(x) < 0, & \text{per ogni } x < c \\ p_2\,f_2(x) - p_1\,f_1(x) > 0, & \text{per ogni } x > c. \end{cases} \qquad (1)$$

Se $p_2\,f_2(x) - p_1\,f_1(x) = 0$, l'osservazione può essere assegnata "a caso" ad una delle due popolazioni.

Ora, essendo note le probabilità a priori p_k e le densità di probabilità $f_k(x)$, si possono calcolare, applicando il *teorema di Bayes*, anche le (densità di)

probabilità a posteriori mediante le

$$p(U_k|X = x) = \frac{p_k f_k(x)}{p_1 f_1(x) + p_2 f_2(x)} 0, \qquad k = 1,2$$

Tramite la (1) si arriva allora alla *regola di Bayes,* mediante la quale l'osservazione $X = x$ viene attribuita a quella popolazione U_k tale che

$$p(U_k|X = x) = \max \qquad (2)$$

Nel caso in cui le probabilità a priori non sono note (caso b), è possibile provare che l'errore di assegnazione è minimo per quel valore c della variabile X tale che

$$\varepsilon_2(c) = \varepsilon_1(c) \qquad (3)$$

ed inoltre si ha:

$$\varepsilon(c) = \frac{n_1}{n}\varepsilon_1(c) + \frac{n_2}{n}\varepsilon_2(c). \qquad (3')$$

In sostanza, con la (3) si postula che un individuo per il quale $X = c$ abbia la stessa probabilità di essere assegnato a U_2 o ad U_1.

Un caso particolarmente significativo si ottiene ipotizzando che la variabile X sia distribuita normalmente nelle due popolazioni. I parametri siano noti o, comunque stimabili mediante le medie e varianze campionarie. Sotto queste condizioni, la (2) richiede la determinazione del

$$\max_{k=1,2} p(U_k|X = x) = \frac{\exp(-1/2\ D_k^2)}{\sum_{h=1}^{2} \exp(-1/2\ D_h^2)}, \qquad (4)$$

essendo $D_k^2 = \ln(s_k^2) + \frac{(x - \bar{x}_k)^2}{s_k^2} - 2\ln(p_k)$; mentre la (3) porta a ricavare il valore soglia come media aritmetica delle medie campionarie ponderate con le deviazioni standard, ossia:

$$c = \frac{\bar{x}_1 s_2 + \bar{x}_2 s_1}{s_1 + s_2}. \qquad (5)$$

Si osservi che in quest'ultimo caso, "in prima approssimazione", come valore soglia può essere preso anche la media aritmetica semplice delle medie

campionarie. Tuttavia in tal modo non si tiene conto della variabilità del fenomeno preso in esame e, quindi, i risultati potrebbero risultare poco attendibili. Solamente se $s_1 \approx s_2$, il valore di c ottenuto come media semplice delle due medie coincide con quello dato dalla (5).

Quanto sopra illustrato è relativo al caso di una sola variabile rilevata su due campioni, ma volendo considerare due o più caratteri presi congiuntamente su ogni unità campionaria è necessario "ricondurre" le variabili stesse ad una sola e, quindi, applicare a questa la metodologia presentata. Si tratta di determinare quella nuova variabile Y, funzione del vettore casuale **X** costituito da q caratteri X_i, che consenta di discriminare gli elementi delle due popolazioni e quindi di assegnare qualunque nuovo elemento che si presenti all'osservazione ad una di esse con la minima probabilità di errore di classificazione. Tale funzione, detta *funzione discriminante*, può essere assunta come combinazione lineare delle variabili X_i, ossia:

$$Y(\mathbf{X}) = \sum_{i=1}^{q} a_i X_i = \mathbf{a}^{\mathrm{I}}\,\mathbf{X} + \beta \qquad (6)$$

con i coefficienti $\mathbf{a}^{\mathrm{I}} = (a_i)$ e la costante β da determinare. In tale formula, come anche in quelle seguenti, l'apice $^{\mathrm{I}}$ denota la trasposta di una matrice.

Da un punto di vista metodologico, estratti i due campioni dalle popolazioni U_1 ed U_2, bisogna anzitutto considerare i vettori delle medie e le matrici di varianza e covarianza (nel seguito v.c.) campionari (rispettivamente $\bar{\mathbf{x}}_k$ e $\mathbf{S}_k$, per k=1,2) e osservare che la matrice **S** di v.c. "totale", ottenuta considerando le unità dei campioni come appartenenti ad un unico campione di numerosità $n_1 + n_2$, è data dalla:

$$\mathbf{S} = \frac{n-2}{n-1}\,\mathbf{W} + \frac{1}{n-1}\,\mathbf{B},$$

essendo **W** e **B**, rispettivamente, le matrici di v.c. "entro" e "tra" i campioni:

$$\mathbf{W} = (w_{ij}) = \sum_{k=1}^{2} \frac{n_k - 1}{n-2}\,\mathbf{S}_k, \quad \mathbf{B} = (b_{ij}) = \frac{n_1 n_2}{n}\left[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right]\left[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right]^{\mathrm{I}}.$$

Si ricerca una variabile Y che assuma valori "molto vicini" per elementi di uno stesso campione e "molto differenti" per individui di campioni diversi. Poiché risulta che la varianza totale di Y è pari a:

$$s^2 = \frac{n-2}{n-1}\,\mathbf{a}^{\mathrm{I}}\mathbf{W}\,\mathbf{a} + \frac{1}{n-1}\,\mathbf{a}^{\mathrm{I}}\mathbf{B}\,\mathbf{a} = \mathbf{a}^{\mathrm{I}}\mathbf{S}\,\mathbf{a},$$

il problema consiste nel massimizzare la varianza "tra" rispetto a quella "entro" i campioni, ossia nel determinare il vettore **a** per il quale si abbia:

$$\max \gamma = \max \frac{\mathbf{a}^{\mathrm{t}} \mathbf{B}\, \mathbf{a}}{\mathbf{a}^{\mathrm{t}} \mathbf{W}\, \mathbf{a}}.$$

E' questo un problema di massimo che, sotto un opportuno *vincolo* per l'unicità di soluzione (ad es. $\mathbf{a}^{\mathrm{t}} \mathbf{S}\, \mathbf{a} = 1$ oppure $\mathbf{a}^{\mathrm{t}} \mathbf{W}\, \mathbf{a} = 1$), può essere affrontato facendo ricorso alla consueta tecnica dei moltiplicatori di Lagrange.
Il problema ha come soluzione un autovettore **a** di $\mathbf{W}^{-1}\mathbf{B}$ corrispondente al massimo autovalore di tale matrice. La soluzione risulta non banale solo se il determinante $|\mathbf{B} - \gamma\, \mathbf{W}| = 0$; osservato inoltre che le matrici **B** e **W** sono simmetriche semidefinite positive e che, supponendo le variabili X_i indipendenti, esse hanno rango, rispettivamente, uguale a 1 e q, si può dire la matrice $\mathbf{W}^{-1}\ \mathbf{B}$ possiede un unico autovalore non nullo δ_0. Per la determinazione dell'autovettore si tratta di risolvere il sistema di equazioni lineari $(\mathbf{W}^{-1}\ \mathbf{B} - \delta_0\, \mathbf{I})\ \mathbf{a} = 0$ nell'incognita **a**, essendo **I** la matrice unitaria qxq. Attraverso alcuni calcoli matriciali si ottiene:

$$\mathbf{a} = \lambda\ \mathbf{W}^{-1} \left[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right],$$

essendo λ un reale positivo determinabile in base al vincolo imposto. La costante β è invece data da: $\beta = -\mathbf{a}^{\mathrm{t}}\, \bar{\mathbf{x}}$, essendo $\bar{\mathbf{x}}$ il vettore "totale" delle medie.

Individuata la variabile Y della (6) si può procedere, attraverso la regola (2) o la (3), all'assegnazione di ogni nuova unità rilevata ad una delle due popolazioni.

Nel caso che il vettore **X** abbia distribuzione multinormale, osservato che in generale la media e varianza di Y nel campione k-mo sono date da $\bar{y}_k = \mathbf{a}^{\mathrm{t}}\, \bar{\mathbf{x}}_k + \beta$ e $s_k^2 = \mathbf{a}^{\mathrm{t}}\, \mathbf{S}_k\, \mathbf{a}$, si possono stimare le probabilità della (4), determinando le quantità D_k^2 con la sostituzione di X con Y, oppure, il valore discriminante c relativo alla variabile Y attraverso la (5). Si fa osservare che le D_k^2 vengono calcolate anche considerando esclusivamente le variabili originali attraverso la $D_k^2 = \ln|\mathbf{S}_k| + \left[\mathbf{x} - \bar{\mathbf{x}}_k\right]^{\mathrm{t}} \mathbf{S}_k^{-1} \left[\mathbf{x} - \bar{\mathbf{x}}_k\right] - 2 \ln(p_k)$ e che diversi software applicativi utilizzano la regola (4) sotto l'assunzione di uguaglianza delle matrici di v.c. delle popolazioni; in tal caso le $\mathbf{S}_h$ presenti nelle D_k^2 possono

essere sostituite dalla matrice **W** e la quantità $\left[\mathbf{x}-\bar{\mathbf{x}}_k\right]' \mathbf{W}_k^{-1}\left[\mathbf{x}-\bar{\mathbf{x}}_k\right]$ è proprio la *distanza di Mahalanobis* del vettore delle osservazioni **x** dal vettore delle medie del campione k-mo.

Una considerazione va fatta in merito alla determinazione degli errori di assegnazione nelle applicazioni. Spesso per la stima di ε_1 ed ε_2 si fa riferimento alle frequenze relative condizionate dei casi, appartenenti ai campioni considerati, erroneamente "riclassificati" attraverso le metodiche discriminanti utilizzate, ovvero

$$\varepsilon_1 \approx \text{Freq (unità classificata in } U_2 \mid \text{unità appartenente ad } U_1\text{)},$$
$$\varepsilon_2 \approx \text{Freq (unità classificata in } U_1 \mid \text{unità appartenente ad } U_2\text{)}; \qquad (7)$$

mentre l'errore complessivo può essere stimato attraverso la (3'). Tale procedimento permette di ottenere una stima accettabile anche se a volte può portare a stime distorte. Se la numerosità complessiva $n = n_1 + n_2$ dei campioni fosse abbastanza grande, i dati campionari potrebbero essere suddivisi in modo "random" in due sottoinsiemi, uno dei quali permetterebbe di ricavare le funzioni discriminanti e l'altro la stima della percentuale di casi erroneamente classificati mediante il modello discriminante costruito con i dati del primo sottoinsieme. In tal modo si otterrebbe una stima degli errori più attendibile; ma per una discussione in merito alla problematica e per ulteriori approfondimenti si rimanda a testi specifici (LACHENBRUCH [3]).

In tutta l'analisi sviluppata finora si è supposto che le due popolazioni esaminate siano distinte rispetto al carattere o ai caratteri rilevati; in realtà va anzitutto verificato se esse presentano differenze statisticamente significative rispetto ai loro valori medi. L'ipotesi nulla H_0 postula l'uguaglianza delle medie della variabile X_i, per $i = 1,2,...,q$, nelle due popolazioni e quella alternativa H_1 la diversità:

$$H_0: \quad \mu_{i1} = \mu_{i2}\,, \qquad H_1: \quad \mu_{i1} \neq \mu_{i2}$$

Per la verifica si fa riferimento alla statistica:

$$F^{(i)} = \frac{b_{ii}}{w_{ii}} \qquad (8)$$

che, sotto H_0, si distribuisce secondo una *F di Snedecor* con $\nu_1 = 1$ e $\nu_2 = n-2$ gradi di libertà (g.d.l.). Un valore elevato di $F^{(i)}$ indica un maggiore contributo alla variabilità totale della X_i da parte della varianza "tra" rispetto a quella "entro" i campioni e, quindi, permette l'attribuzione di gran parte della variabilità totale alla differenza tra le medie dei campioni e il rifiuto conseguente dell'ipotesi nulla. In pratica, l'ipotesi di uguaglianza tra le medie è rifiutata, ad un dato livello di significatività α, se il valore di $F^{(i)}$

ottenuto sui dati campionari a disposizione risulta maggiore della F "teorica" $F_{\alpha,1,n-2}$.

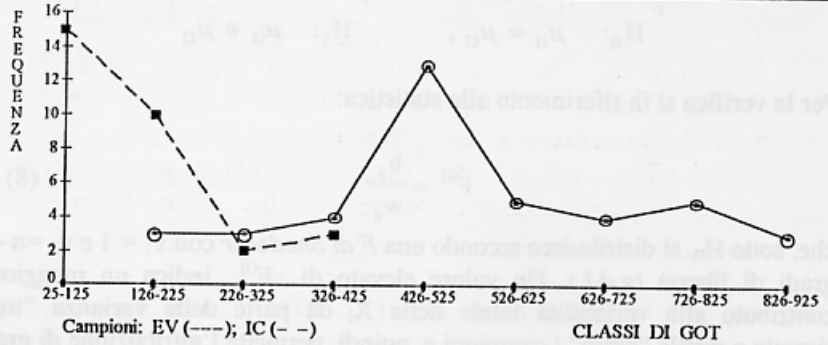
Un altro problema è quello della verifica dell'uguaglianza delle matrici di v.c. delle popolazioni. A tale scopo si utilizza il *criterio di Box* (BOX [1]); ma per una trattazione dell'argomento e di tutto quanto connesso con la verifica di ipotesi si rimanda a testi metodologici e applicativi (DELVECCHIO [2], MC LACHALAN [4], TAMPIERI [5]).

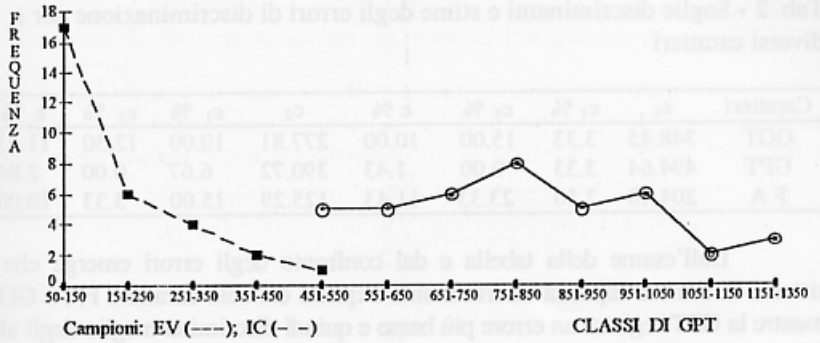
RISULTATI

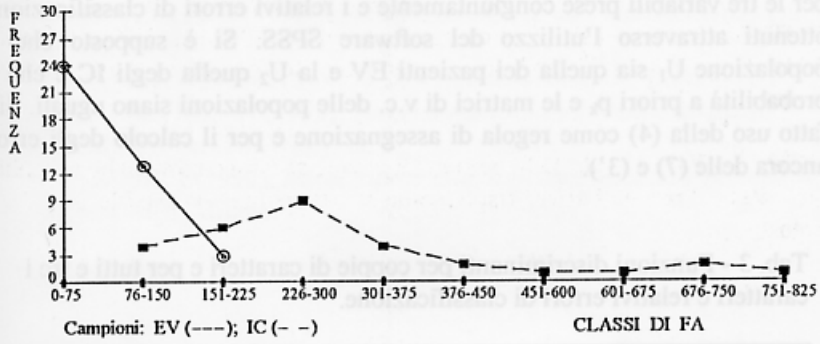
Nella Tab. 1 viene riportata una sintesi statistica dei valori delle attività GOT, GPT e FA del siero nei campioni di pazienti affetti da EV e IC, mentre nelle Figg. seguenti vengono diagrammate le rispettive distribuzioni di frequenza.

Tab. 1 - Valori medi, D.S. e range di alcune attività enzimatiche eseguite su 40 e 30 pazienti affetti rispettivamente da epatite virale e ittero colostatico

	EPATITE VIRALE			ITTERO COLOSTATICO		
Variabile	media	D.S.	range	media	D.S.	range
G.O.T	539.90	192.44	168-923	157.00	88.70	50-420
G.P.T	821.87	218.82	503-1258	167.40	113.34	59-503
F. A	81.85	40.32	32-221	327.10	187.31	98-820







Dall'esame della tabella e dei grafici appare evidente come l'accertamento di laboratorio che, singolarmente considerato, registra la maggiore efficacia discriminante sia la GPT.

I coefficienti di correlazione positivi e statisticamente significativi sono emersi tra i caratteri GOT e GPT sia nella casistica di pazienti di EV ($r = 0.76$; prob. < 0.01) che in quella di IC ($r = 0.95$, prob. < 0.001).

Le popolazioni risultano distinte rispetto ai caratteri in quanto i valori delle $F^{(i)}$ ottenuti tramite la (8) sono "elevati" ($F^{(GOT)} = 102.19$; $F^{(GPT)} = 222.91$; $F^{(FA)} = 64.87$ e, in tutti i casi, prob. ≈ 0.0000).

Nella Tab.2 vengono riportati, per ogni singolo carattere preso in esame, i valori soglia (c_1 ottenuto come media semplice delle medie campionarie e c_2 come media delle medie ponderate con le varianze) e le stime dei relativi errori di discriminazione, determinati rispettivamente attraverso le formule (5), (7) e (3'). Si fa osservare che per i caratteri GOT e GPT la popolazione U_1 è quella dei pazienti IC e la U_2 quella dei pazienti EV, mentre per FA la situazione è invertita.

Tab. 2 - Soglie discriminanti e stime degli errori di discriminazione per i diversi caratteri

Caratteri	c_1	ε_1 %	ε_2 %	ε %	c_2	ε_1 %	ε_2 %	ε %
GOT	348.45	3.33	15.00	10.00	277.81	10.00	12.50	11.43
GPT	494.64	3.33	0.00	1.43	390.72	6.67	0.00	2.86
F A	204.48	2.50	23.33	11.43	125.29	15.00	3.33	10.00

Dall'esame della tabella e dal confronto degli errori emerge che il sistema di minore capacità discriminante è quello del solo carattere FA e GOT, mentre la GPT registra un errore più basso e quindi discrimina meglio degli altri due caratteri.

La Tab. 3 riporta le funzioni discriminanti per coppie di variabili e per le tre variabili prese congiuntamente e i relativi errori di classificazione, ottenuti attraverso l'utilizzo del software SPSS. Si è supposto che la popolazione U_1 sia quella dei pazienti EV e la U_2 quella degli IC e che le probabilità a priori p_k e le matrici di v.c. delle popolazioni siano uguali. Si è fatto uso della (4) come regola di assegnazione e per il calcolo degli errori ancora delle (7) e (3').

Tab. 3 - Funzioni discriminanti per coppie di caratteri e per tutti e tre i caratteri e relativi errori di classificazione.

Caratteri	ε_1	ε_2	ε
GOT–GPT *	5.00 %	0.00 %	2.86 %
GOT–FA **	10.00 %	3.33 %	7.14 %
GPT– FA ***	0.00 %	1.67 %	0.71 %
GOT–GPT–FA ****	0.00 %	0.00 %	0.00 %

Funzioni discriminanti:

* $Y = -0.0018\ GOT + 0.0067\ GPT - 2.9225;$

** $Y = 0.0049\ GOT - 0.0047\ FA - 0.9636;$

*** $Y = 0.0048\ GPT - 0.0035\ FA - 1.9567;$

**** $Y = -0.0019\ GOT + 0.0061\ GPT - 0.0035\ FA - 1.8901.$

Nella Tab. 3 si osserva che l'analisi congiunta di due caratteri consente una notevole riduzione della percentuale di errore; solamente la coppia GOT-FA (variabili che singolarmente presentano le maggiori percentuali di errore) fa registrare un errore del 7.14 %. Mentre il modello che comprende le GPT-FA già annulla praticamente l'errore di classificazione. La funzione che comprende tutti i caratteri presi in esame (GOT-GPT-FA) è caratterizzata dal 100 % di casi correttamente classificati.

In conclusione lo studio comparativo dei diversi modelli conferma l'importanza dell'applicazione dell'Analisi Discriminante alla diagnostica. Lo studio ha consentito di individuare i caratteri che danno il maggior contributo

alla diagnosi differenziale. In estrema sintesi è stato ottenuto che la migliore discriminazione dei casi si ottiene dai modelli che comprendono GPT-FA e GOT-GPT-FA.

BIBLIOGRAFIA

1. G.E.P BOX (1949), *A general distribution theory for a class of likelihood criteria*, Biometrika, 36, 317-346.
2. F. DELVECCHIO (1992), *Analisi statistica di dati multidimensionali*, Cacucci, Bari.
3. P.A. LACHENBRUCH (1975), *Discriminant Analysis*, Hafner, N.Y.
4. G. MC LACHALAN (1992), *Discriminant Analysis and Statistical Pattern Recognition,* Wiley, N.Y.
5. A. TAMPIERI (1996), *Introduzione alla Statistica Medica e Biometria*, McGraw-Hill, Milano.