dal gruppo 1

- se $y^* > \frac{\bar{y}_1 + \bar{y}_2}{2} \Rightarrow \underline{x}^*$ proviene dal gruppo 2
- se $y^* = \frac{\bar{y}_1 + \bar{y}_2}{2} \Rightarrow$??

β) Sia $\bar{y}_1 > \bar{y}_2$ allora

- se $y^* < \frac{\bar{y}_1 + \bar{y}_2}{2} \Rightarrow \underline{x}^*$ proviene dal gruppo 2
- se $y^* > \frac{\bar{y}_1 + \bar{y}_2}{2} \Rightarrow \underline{x}^*$ proviene dal gruppo 1
- se $y^* = \frac{\bar{y}_1 + \bar{y}_2}{2} \Rightarrow$??

Individuazione delle variabili che maggiormente contribuiscono a spiegare l'appartenenza ad una particolare sottopopolazione (CENNI)

→ valutazione diretta dell'ordine di grandezza dei coefficienti a_i

→ calcolo della correlazione tra i punteggi discriminanti (SCORES) e le q variabili

VALUTAZIONE DELLA BONTÀ DELLA

VALUTAZIONE DELLA BONTÀ DELLA FUNZIONE DISCRIMINANTE come CRITERIO DI CLASSIFICAZIONE.

Calcolo del tasso di corretta riclassificazione

Riclassifico tutte le unità statistiche confrontando il loro punteggio discriminante con il CUT OFF ottenendo la seguente tabella di contingenza:

g.r. → gruppo in cui si viene riclassificati

g.a. → gruppo di appartenenza

g.a. \ g.r.	I°	II°	
I°	n_{11}	n_{12}	n_1
II°	n_{21}	n_{22}	n_2
			n

$$TCR = \frac{n^\circ \text{ di correttamente riclassificati}}{n}$$

$$= \frac{n_{11} + n_{22}}{n}$$

OSSERVAZIONE: un limite del TCR sta nel fatto che esso è valutato sulle unità statistiche che hanno contribuito alla unità statistiche che hanno contribuito alla definizione della funzione discriminante. Per questa ragione va interpretato "in negativo" nel senso che un elevato TCR non assicura la bontà della regola di classificazione mentre un basso TCR ne evidenzia lo scarso valore.

Esempio Numerico

Training set: 10 aziende a cui una banca ha erogato, in passato un prestito. 5 aziende hanno onorato la loro obbligazione (gruppo 1) mentre le restanti sono risultate insolventi (gruppo 2). Sulle 10 aziende si sono rilevati i seguenti indici di bilancio:

$X_1 = ROE =$ "Return on equity"

↳ redditività del patrimonio

$$X_2 = \frac{\text{ONERI FINANZIARI}}{\text{RICAVI}}$$

↳ indice di indebitamento

Matrice dei dati

$$X = \left[\begin{array}{cc} 0,02 & 0,8 \\ 0,2 & 1,2 \\ 0,14 & 0,05 \\ -0,13 & 0,55 \\ 0,12 & 0,36 \\ \hdashline -0,02 & 0,75 \\ 0,05 & 1,1 \\ 0,01 & 0,95 \\ -0,3 & 0,99 \\ -0,05 & 1,36 \end{array}\right] \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} \text{gruppo} \\ \left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} \text{gruppo} \end{array}$$

$$\underline{y} = X\underline{a} \qquad \text{dove} \qquad \underline{a} = S^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

Calcolo della matrice S di varianze - covarianze e di $(\underline{\mu}_1 - \underline{\mu}_2)$

μ_{11} = Media di X_1 nel gruppo 1 =

$$= \frac{0,02 + 0,20 + 0,14 - 0,13 + 0,12}{5} = 0,07$$

μ_{12} = Media di X_2 nel gruppo 2 =

$$= 0,592$$

$$\underline{\mu}_1 = \begin{bmatrix} 0,07 \\ 0,592 \end{bmatrix}$$

In modo del tutto analogo si ottiene

$$\underline{\mu}_2 = \begin{bmatrix} -0,062 \\ 1,03 \end{bmatrix}$$

$$\Rightarrow (\underline{\mu}_1 - \underline{\mu}_2) = \begin{bmatrix} 0,132 \\ -0,438 \end{bmatrix}$$

Per calcolare S serve anche il vettore delle medie "generali" $\underline{\mu}$

$$\underline{\mu} = \frac{1}{10}\left[5 \cdot \underline{\mu}_1 + 5\,\underline{\mu}_2\right] = \begin{bmatrix} 0{,}004 \\ 0{,}811 \end{bmatrix}$$

(μ_{x_2}) media di X_2 sulle 10 aziende

media di X_1 sulle 10 aziende (μ_{x_1})

$$S = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma^2_{x_2} \end{bmatrix}$$

$$\sigma^2_{x_1} = \frac{1}{10}\sum_{i=1}^{10} x_{i1}^2 - \mu_{x_1}^2$$

$$= 0{,}01868 - 0{,}004^2 = 0{,}018664$$

$$\sigma^2_{x_2} = \frac{1}{10}\sum_{i=1}^{10} x_{i2}^2 - \mu_{x_2}^2 = 0{,}144209$$

$$\sigma_{x_1x_2} = \frac{1}{10}\sum_{i=1}^{10} x_{i1}\,x_{i2} - \mu_{x_1}\,\mu_{x_2}$$

$$= -0{,}011324$$

$$S = \begin{bmatrix} 0,018664 & -0,011324 \\ -0,011324 & 0,144209 \end{bmatrix}$$

$$S^{-1} = \frac{1}{\det(S)} \begin{bmatrix} 0,144209 & 0,011324 \\ 0,011324 & 0,018664 \end{bmatrix}$$

$$\det(S) = 0,018664 \cdot 0,144209 - 0,011324^2$$

$$= 0,0025633$$

$$S^{-1} = \begin{bmatrix} 56,259475 & 4,417771 \\ 4,417771 & 7,281285 \end{bmatrix}$$

$$a = S^{-1}(\underline{\mu}_1 - \mu_2)$$

$$= \begin{bmatrix} 56,26 & 4,42 \\ 4,42 & 7,28 \end{bmatrix} \begin{bmatrix} 0,004 \\ 0,811 \end{bmatrix}$$

$$= \begin{bmatrix} 5,4912 \\ -0,26060 \end{bmatrix}$$

$$a = \begin{bmatrix} 5,4912 \\ -0,26060 \end{bmatrix}$$

(a può essere eventualmente normalizzato)

$$a^* = \frac{a}{\sqrt{a^T a}} = \frac{1}{6,078}\, a$$

$$= \begin{bmatrix} 0,9034 \\ -0,4287 \end{bmatrix}$$

Calcolo dei punteggi discriminanti

$$y = X a^*$$

$$= \begin{bmatrix} -0,32493 \\ -0,33381 \\ 0,10504 \\ -0,35326 \\ -0,04594 \\ -0,33963 \\ -0,42645 \\ -0,39829 \\ -0,69549 \\ -0,62827 \end{bmatrix}$$

$$\bar{y}_1 = -0,19058$$

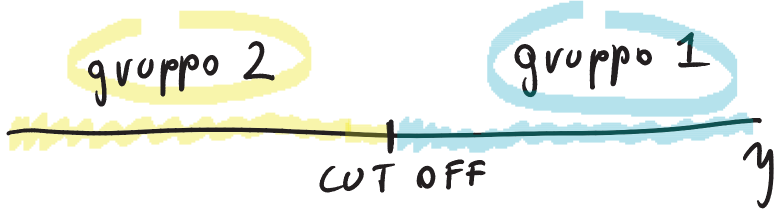
$$\bar{y}_2 = -0,49762$$

g.d. \ g.v.	1	2	
1	4	1	5
2	1	4	5
			10

cut off = -0,19058 - 0,49762

$$\text{cut off} = \frac{-0{,}19038 - 0{,}49762}{2}$$

$$= -0{,}34410$$



$$TCR = \frac{8}{10} \Rightarrow 80\% \text{ di unità}$$

statistiche (aziende) correttamente riclassificate.

$\Rightarrow$ Si presenta un nuovo cliente che richiede un prestito presentando dati di bilancio che portano a:

$$x_1^* = 0{,}20 \qquad x_2^* = 1{,}50$$

Il cliente è affidabile o no?

Sia $\underline{x}^* = [\, x_1^* \; ; \; x_2^* \,]$.

Calcolo il punteggio discriminante di $\underline{x}^*$:

$$y^* = \underline{x}^* \, \underline{a}^* = [\,0{,}2 \quad 1{,}5\,] \begin{bmatrix} 0{,}9034 \\ -0{,}4287 \end{bmatrix}$$

$$= -0,4624$$

$y^* <$ CUT-OFF $\Rightarrow$ la nuova azienda viene classificata tra le insolventi.

Cosa avremmo ottenuto se avessimo calcolato a come $a = S_W^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$?

$$S_1 = \begin{bmatrix} 0,01336 & 0,0055 \\ 0,0055 & 0,152456 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0,015256 & 0,00076 \\ 0,00076 & 0,04004 \end{bmatrix}$$

$$S_W = \begin{bmatrix} 0,014308 & 0,003130 \\ 0,003130 & 0,096248 \end{bmatrix}$$

$$S_W^{-1} = \begin{bmatrix} 70,391742 & -2,28915 \\ -2,28915 & 10,46427 \end{bmatrix}$$

$$a = S_w^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = \begin{bmatrix} 10{,}29435 \\ -4{,}885518 \end{bmatrix}$$??

$$a^* = \frac{a}{\sqrt{a^T a}} = \begin{bmatrix} 0{,}90343 \\ -0{,}42875 \end{bmatrix}$$