come punti in uno spazio n-dimensionale che chiameremo *spazio degli oggetti*. Per esempio la matrice di correlazione ha una naturale interpretazione nello spazio degli oggetti della matrice centrata $\mathbf{Y} = \mathbf{H}\underset{\sim}{\mathbf{X}}$. La correlazione è proprio il coseno dell'angolo θ_{ij} all'origine fra le colonne i e j:

$$\cos\theta_{ij} \;=\; \frac{\mathbf{y}'_{(i)}\mathbf{y}_{(j)}}{\|\mathbf{y}_i\|\,\|\mathbf{y}_j\|} \;=\; \frac{s_{ij}}{s_i\,s_j} \;=\; r_{ij}\ .$$

Si noti che i coefficienti di correlazione sono misure di *similarità* in quanto i loro valori sono grandi quando le variabili sono linearmente dipendenti.

Nel secondo caso, le n righe possono essere viste come n punti in uno spazio p-dimensionale (*spazio delle variabili*). Un modo naturale di confrontare due righe $\mathbf{x}_h$ e $\mathbf{x}_k$ è quello di valutare la distanza euclidea:

$$\|\mathbf{x}_h - \mathbf{x}_k\|^2 \;=\; (\mathbf{x}_h - \mathbf{x}_k)'(\mathbf{x}_h - \mathbf{x}_k)\ .$$

Un modo alternativo di procedere è quello di trasformare i dati mediante una delle trasformazioni lineari viste nei paragrafi 1.6.1 e 1.6.2 e considerare le distanze rispetto a tali trasformazione delle righe. Tali distanze giocano un ruolo importante nell'analisi dei gruppi. La più importante di tali distanze è la distanza di Mahalanobis

$$\delta_M(\mathbf{z}_h, \mathbf{z}_k) \;=\; \|\mathbf{z}_h - \mathbf{z}_k\|^2 \;=\; (\mathbf{x}_h - \mathbf{x}_k)'\mathbf{S}^{-1}(\mathbf{x}_h - \mathbf{x}_k)\ .$$

Tale distanza svolge un ruolo importante nell'analisi discriminante. Si noti infine che tale distanza può essere definita anche considerando la matrice $\mathbf{S}_u$ piuttosto che la matrice $\mathbf{S}$.

1.8 Matrici di distanze

1.8.1 Distanze, indici di dissimilarità, indici di similarità

Affrontiamo il problema della misurazione della "prossimità" fra due unità statistiche ω_i e ω_j a cui corrispondono due diverse righe $\mathbf{x}_i$' e $\mathbf{x}'_j$ della matrice dei dati. Si supponga, ad esempio, di aver rilevato per n autovetture p variabili (cilindrata, prezzo, velocità massima, etc.). La conoscenza degli indici di prossimità per ciascuna delle possibili coppie consente di individuare quelle tra loro più simili (meno diverse). Un altro esempio è fornito dai comuni di una regione, per i quali si possono costruire molti indicatori demografici, economici e sociali, in base ai dati del censimento della popolazione e di quello dell'industria e dei servizi: gli indici di prossimità consentono di "misurare" la diversità degli stessi, sotto gli aspetti via via considerati.

In una matrice di dati, si possono ovviamente considerare anche caratteri qualitativi, le cui modalità sono tradotte da opportuni codici numerici. Un esempio è fornito dallo spoglio di n questionari di un'inchiesta che preveda p domande con risposte chiuse di tipo qualitativo (es. professione, sesso, tipo di letture, etc.) e di tipo quantitativo (età, ammontare per spes domestiche, etc.). Per due generici rispondenti all'inchiesta, uò essere interessante valutare il grado di prossimità o di diversità fra i due insiemi di risposte.

Gli indici di prossimità vengono abitualmente distinti a seconda che si considerino fenomeni quantitativi o fenomeni qualitativi. Con riferimento al primo caso verranno considerati le distanze, gli indici di distanza e gli indici di dissimilarità, che corrispondono a

famiglie via via più ampie di indici, in cui ciascuna classe comprende le precedenti; per i caratteri qualitativi verranno presentati gli indici di similarità. Infine verrà affrontato il caso misto, in cui compaiono sia caratteri quantitativi e qualitativi. Per ulteriori approfondimenti, vedi [Zan00].

Definizione 1.3 Si chiama *distanza* una funzione $d : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ che gode delle seguenti proprietà:

a) *non negatività*: $d(\mathbf{x}, \mathbf{y}) =\geq 0$ per ogni $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$

b) *identità*: $d(\mathbf{x}, \mathbf{y}) = 0$ se e solo se $\mathbf{x} \equiv \mathbf{y}$;

c) *simmetria*: $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ per ogni $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$;

d) *disuguaglianza triangolare*: $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ per ogni $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^p$.

La coppia $(\mathbb{R}^p, d)$ viene chiamata *spazio metrico*.

Nel seguito considereremo distanze su spazi $\mathbb{R}^p$, e per $1 \leq q < \infty$ si introduce la norma:

$$\|\mathbf{x}\|_q := \left(\sum_{i=1}^{p} |x_i|^q \right)^{1/q} .$$

Un'altra norma è $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq p} |x_i|$ e si può dimostrare che $\lim_{q \to \infty} \|\mathbf{x}\|_q = \|\mathbf{x}\|_\infty$ per ogni $\mathbf{x} \in \mathbb{R}^p$, fatto che rende ragione della notazione $\|\mathbf{x}\|_\infty$.

La distanza di Minkowski. Nelle applicazioni notevole importanza presenta la *distanza di Minkowski* di ordine q definita su $\mathbb{R}^p \times \mathbb{R}^p$ da:

$$d_q(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_q = \left[\sum_{h=1}^{p} |x_h - y_h|^q \right]^{1/q} . \tag{1.33}$$

Innanzitutto dimostriamo che la distanza di Minkowski soddisfa le assunzioni della Definizione 1.3. Le proprietà *a*), *b*) e *c*) sono ovvie; per quanto riguarda la disuguaglianza triangolare si dimostra - tramite la disuguaglianza di Minkowski di cui viene omessa la dimostrazione - che per ogni $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^p$ si ha:

$$\|\mathbf{x} - \mathbf{y}\|_q \leq \|\mathbf{x} - \mathbf{z}\|_q + \|\mathbf{z} - \mathbf{y}\|_q .$$

In corrispondenza di particolari valori di q si ottengono alcuni casi notevoli.

La distanza di Manhattan ($q = 1$). Per $q = 1$ nella (1.33) si ottiene:

$$d_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1 = \sum_{h=1}^{p} |x_h - y_h| . \tag{1.34}$$

che viene chiamata metrica di *Manhattan* o metrica della *città a blocchi* in quanto rappresenta la distanza che deve percorrere un individuo che di muove in una città con strade parallele e perpendicolari.

Distanza Euclidea ($q = 2$). Per $q = 2$ nella (1.33) si ottiene:

$$d_2(\mathbf{x}, \mathbf{y}) \;=\; \|\mathbf{x} - \mathbf{y}\|_2 \;=\; \sqrt{\sum_{h=1}^{p} |x_h - y_h|^2} \; . \tag{1.35}$$

che costituisce la nota distanza euclidea.

Distanza di Lagrange ($q \to \infty$). Per $q \to \infty$ nella (1.33) si ottiene:

$$d_\infty(\mathbf{x}, \mathbf{y}) \;=\; \lim_{q \to \infty} d_q(\mathbf{x}, \mathbf{y}) \;=\; \max_{h=1,\ldots,n} |x_h - y_h| \tag{1.36}$$

che viene chiamata distanza di *Lagrange* o metrica L_∞ o *norma sup*. Essa discende dalla nota proprietà della somma di potenze:

$$\lim_{q \to \infty} \left(\sum_{h=1}^{p} |a_h|^q \right)^{1/q} \;=\; \max\{a_1, a_2, \ldots, a_n\} \; .$$

La distanza euclidea, e quindi la relativa geometria, hanno trovato molte applicazioni nel campo della fisica, della tecnologia etc. In molte situazioni di ricerca statistica, la distanza della città a blocchi è da preferirsi a quella euclidea; ciò in particolar modo nel campo socio-economico. La distanza euclidea tende a dare maggiore importanza alle differenze tra coordinate più grandi mentre quella della città a blocchi pone sullo stesso livello tutte le differenze.

Ogni scelta metrica impone, quindi, una diversa geometria di riferimento. In particolare, il modello euclideo è uno dei tanti possibili; esso non è privilegiato in statistica rispetto ad altri.

1.8.2 Proprietà delle distanze di Minkowski

La distanze di Minkowski gode di alcune importanti proprietà; qui ci limiteremo a dare solo quelle più importanti ai fini statistici.

Proprietà di monotonia con q. Si dimostra che per la metrica di Minkowski valgono le seguenti relazioni:

$$d_1(\mathbf{x}, \mathbf{y}) \geq d_2(\mathbf{x}, \mathbf{y}) \geq \cdots \geq d_\infty(\mathbf{x}, \mathbf{y}) \; ,$$

cioè, assegnati $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, la loro distanza diminuisce al crescere di q.

Invarianza per traslazione. La distanza di Minkowski è invariante per traslazione delle variabili:

$$d_q(\mathbf{x} + \mathbf{c}, \mathbf{y} + \mathbf{c}) \;=\; d_q(\mathbf{x}, \mathbf{y})$$

dove $\mathbf{c} \in \mathbb{R}^q$ è un vettore costante. Infatti, si ha:

$$d_q(\mathbf{x} + \mathbf{c}, \mathbf{y} + \mathbf{c}) = \|\mathbf{x} + \mathbf{c} - (\mathbf{y} + \mathbf{c})\|_q = \|\mathbf{x} - \mathbf{y}\|_q = d_q(\mathbf{x}, \mathbf{y}) \; .$$

Questa proprietà è molto importante in statistica, poichè consente di affermare, fra l'altro, che la distanza rimane invariata quando essa viene calcolata, anzichè sui valori originari delle variabili, sui rispettivi scostamenti dalla media. In particolare, se $\mathbf{c} = -\boldsymbol{\mu}$, si ha:

$$d_q(\mathbf{x} - \boldsymbol{\mu}, \mathbf{y} - \boldsymbol{\mu}) = d_q(\mathbf{x}, \mathbf{y}) \, ,$$

cioè la distanza rimane invariata quando essa viene calcolata, anzichè sui valori originari, sui rispettivi scarti dalla media.

Invarianza della distanza euclidea per trasformazioni ortogonali. La distanza euclidea è invariante per trasformazioni ortogonali[13] (rotazioni) delle variabili, cioè

$$d_2(\mathbf{Tx}, \mathbf{Ty}) \;\; = \;\; d_2(\mathbf{x}, \mathbf{y})$$

dove $\mathbf{T}$ è una matrice $p \times p$ ortogonale, cioè tale che $\mathbf{T}'\mathbf{T} = \mathbf{I}$, tenendo conto che $d_2^2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2 = (\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})$. Si ha infatti:

$$\begin{aligned} d_2^2(\mathbf{Tx}, \mathbf{Ty}) &= \|\mathbf{Tx} - \mathbf{Ty}\|_2^2 = \|\mathbf{T}(\mathbf{x} - \mathbf{y})\|_2^2 = [\mathbf{T}(\mathbf{x} - \mathbf{y})]'[\mathbf{T}(\mathbf{x} - \mathbf{y})] \\ &= (\mathbf{x} - \mathbf{y})'\mathbf{T}'\mathbf{T}(\mathbf{x} - \mathbf{y}) = (\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2 = d_2^2(\mathbf{x}, \mathbf{y}) \, . \end{aligned}$$

Ne segue che le distanze euclidee tra i punti in $\mathbb{R}^p$ non mutano quando si effettua sia una traslazione degli assi (proprietà precedente), sia una rotazione degli stessi.

1.8.3 Matrici di distanze

Nel caso di una matrice di dati con variabili tutte quantitative, la distanza fra due generiche unità statistiche corrisponde ad una classe particolare e molto importante calcolata sui vettori riga $\mathbf{x}_i$ e $\mathbf{x}_j$ e sarà indicata con $d(\mathbf{x}_i, \mathbf{x}_j) = d_{ij}$.

Con riferimento a n unità statistiche , calcolando la distanza d_{ij} fra tutte le possibili coppie di elementi si ottiene la *matrice delle distanze*, che indichiamo con **D**:

$$\mathbf{D} = \begin{pmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ & 0 & d_{23} & \cdots & d_{2n} \\ & & & \ddots & \vdots \\ & & & & 0 \end{pmatrix}$$

[13] **Definizione.** Una matrice quadrata **A** si dice *ortogonale* se $\mathbf{AA}' = \mathbf{I}$.
Seguono immediatamente le seguenti propriet`a delle matrici ortogonali:

1. $\mathbf{A}^{-1} = \mathbf{A}'$,
2. $|\mathbf{A}| = \pm 1$,
3. Siano $\mathbf{a}_i, \mathbf{a}_j$ le righe i, j della matrice **A**, si ha allora $\mathbf{a}_i'\mathbf{a}_j = 1$ se $i = j$ e $\mathbf{a}_i'\mathbf{a}_j = 0$ se $i \neq j$; analogamente, se indichiamo con $\mathbf{a}_{(i)}, \mathbf{a}_{(j)}$ le colonne i, j della matrice **A**, si ha $\mathbf{a}_{(i)}\mathbf{a}_{(j)} = 1$ se $i = j$ e $\mathbf{a}_{(i)}\mathbf{a}_{(j)} = 0$ se $i \neq j$,
4. la matrice $\mathbf{C} = \mathbf{AB}$ `e ortogonale se le matrici **A** e **B** sono ortogonali.

Si noti che tutte queste propriet`a derivano direttamente dalla defi nizione $\mathbf{AA} = \mathbf{I}$. Il risultato 4. afferma che la somma dei quadrati degli elementi in ciascuna riga (colonna) `e uguale a uno mentre la somma dei prodotti degli elementi di righe (colonne) diverse `e nulla.

Le matrici ortogonali possono essere utilizzate per rappresentare un cambio di base o delle rotazioni.

Table 1.1: *Ore di trasmissione nel 1997 delle principali reti televisive per tipo di programma*

Rete	*Film*	*Telefilm*	*Varietà*	*Notiziari*	*Altro*
RaiUno	1158	1280	1577	1703	3042
RaiDue	731	1366	1280	1019	4364
RaiTre	1454	675	937	1618	4076
Rete4	2053	1289	489	1410	3519
Canale5	582	1193	2166	3372	1447
Italia 1	1167	3119	795	1261	2418

Table 1.2: *Matrice di distanze euclidee fra le reti televisive per i dati in Tabella 1.1*

d_2	*RaiUno*	*RaiDue*	*RaiTre*	*Rete4*	*Canale5*	*Italia1*
RaiUno	0.000	863.591	932.988	1438.992	1863.284	2046.678
RaiDue	863.591	0.000	1215.179	1591.281	2524.626	1885.973
RaiTre	932.988	1215.179	0.000	989.831	2369.740	2490.606
Rete4	1438.992	1591.281	989.831	0.000	2972.344	2061.488
Canale5	1863.284	2524.626	2369.740	2972.344	0.000	3222.990
Italia1	2046.678	1885.973	2490.606	2061.488	3222.990	0.000

Per definizione di distanza, tale matrice è simmetrica e presenta valori nulli sulla diagonale principale. Inoltre **D** è semidefinita positiva.

Esempio 1.4 (Analisi della distanza fra la programmazione di alcune reti televisive.) In Tabella 1.1 vengono riportate le ore di trasmissione delle principali reti televisive italiane per tipo di programma. Si vuole valutare la diversità esistente fra le reti con riferimento alla tipologia di programmazione.

Consideriamo le distanze fra le varie reti, con riferimento alle prime quattro variabili, escludendo l'ultima che – essendo residuale – può avere contenuti molto diversi. Consideriamo la distanza euclidea e quella di Manhattan. Da tali matrici si evince che le due reti con tipo di programma più simile (cioè con minore distanza) risultano, con entrambi i criteri, RaiUno e RaiDue, mentre quelle maggiormente diverse sono Canale5 e Italia1 sulla base delle variabili: Film, Telefilm, Varietà, Notiziari. ♣

1.8.4 Indici di distanza ed indici di dissimilarità

Rinunciando ad alcune delle quattro proprietà che definiscono una distanza si ottengono famiglie più ampie di indici di prossimità, che risultano però meno rigorosi ed in talune circostanza possono dare origine ad incongruenze.

Definizione 1.5 Si dice *indice di distanza* fra due vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ una funzione che soddisfa le proprietà di non negatività, identità e simmetria.

Table 1.3: *Matrice di distanze di Manhattan fra le reti televisive per i dati in Tabella 1.1*

d_1	*RaiUno*	*RaiDue*	*RaiTre*	*Rete4*	*Canale5*	*Italia1*
RaiUno	0.000	1494.000	1626.000	2285.000	2921.000	3072.000
RaiDue	1494.000	0.000	2356.000	2581.000	3561.000	2916.000
RaiTre	1626.000	2356.000	0.000	1869.000	4373.000	3230.000
Rete4	2285.000	2581.000	1869.000	0.000	5206.000	3171.000
Canale5	2921.000	3561.000	4373.000	5206.000	0.000	5993.000
Italia1	3072.000	2916.000	3230.000	3171.000	5993.000	0.000

L'esempio più noto di indice di distanza è il quadrato della distanza euclidea:

$$d_2^2(\mathbf{x}, \mathbf{y}) \;=\; \|\mathbf{x} - \mathbf{y}\|_2^2 \;=\; \sum_{i=1}^{p} (x_i - y_i)^2 \,. \tag{1.37}$$

Tale indice non soddisfa la disuguaglianza triangolare, come si evince considerando il seguente semplice esempio. Siano:

$$\mathbf{x} = (10, 5)' \qquad \mathbf{y} = (11, 7)' \qquad \mathbf{z} = (13, 9)'$$

per i quali risulta:

$$d_2^2(\mathbf{x}, \mathbf{y}) = 5 \qquad d_2^2(\mathbf{x}, \mathbf{z}) = 25 \qquad d_2^2(\mathbf{y}, \mathbf{z}) = 8$$

e pertanto risulta: $d_2^2(\mathbf{x}, \mathbf{z}) > d_2^2(\mathbf{x}, \mathbf{y}) + d_2^2(\mathbf{y}, \mathbf{z})$.

L'indice di distanza definito come quadrato della distanza euclidea gode però dell'importante proprietà di addività. Considerati $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, allora la somma dell'indice calcolato su un sottoinsieme di p_1 variabili (componenti) di $\mathbf{x}, \mathbf{y}$ e del valore calcolato sulle restanti p_2 componenti (variabili), è uguale al valore dell'indice calcolato direttamente su tutte le p variabili:

$$\sum_{h=1}^{p_1} (x_h - y_h)^2 + \sum_{h=p_1+1}^{p} (x_h - y_h)^2 \;=\; \sum_{h=1}^{p} (x_h - y_h)^2 \,,$$

con $p_1 + p_2 = p$. Si verifica immediatamente che tale proprietà non è soddisfatta dalla distanza euclidea. Un'ulteriore categoria di indici è la seguente.

Definizione 1.6 Si dice *indice di dissimilarità* fra $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ una funzione che soddisfa le proprietà di non negatività e di simmetria oltre alla proprietà:

$$\mathbf{x} = \mathbf{y} \;\Rightarrow\; d(\mathbf{x}, \mathbf{y}) = 0 \,.$$

Si noti che quest'ultima proprietà è più debole di quella di identità.

La classe degli indici di dissimilarità è più ampia di quella degli indici di distanza; in particolare un indice di dissimilarità può risultare nullo anche quando $\mathbf{x} \neq \mathbf{y}$.

1.8.5 Indici di similarità

Sino ad ora abbiamo considerato indici di prossimità applicabili nel caso in cui gli elementi considerati siano rappresentati da vettori p-dimensionali, corrispondenti ad altre variabili. Consideriamo ora un'altra classe di indici, applicabili anche a fenomeni qualitativi.

Definizione 1.7 Dato un insieme finito di elementi E, si definisce *indice di similarità* un'applicazione $S : E \times E \to \mathbb{R}$ che soddisfa le seguenti tre proprietà:

1. *non negatività*: $S(e_i, e_j) \geq 0$ per ogni $e_i, e_j \in E$,
2. *normalizzazione*: $S(e_i, e_i) = 1$ per ogni $e_i \in E$,
3. *simmetria*: $S(e_i, e_j) = S(e_j, e_i)$ per ogni $e_i, e_j \in E$.

Si noti che un indice di similarità è definito con riferimento agli elementi di un insieme (di unità statistiche) anzichè ai vettori di $\mathbb{R}^p$. Inoltre un indice di similarità assume valori nell'intervallo chiuso $[0, 1]$, mentre una distanza può presentare qualunque valore non negativo.

1.8.6 Indici di similarità per fenomeni dicotomici

Un'importante applicazione degli indici di similarità riguarda i fenomeni dicotomici cioè quelli che presentano solo due modalità. Ad esempio, per n esercizi alberghieri si rileva la presenza o l'assenza di un certo numero p di servizi ed attrezzature (piscina, salone per conferenze, frigorifero in camera, televisore in camera, etc.). Un indice di similarità si prefigge di valutare la prossimità tra coppie di unità statistiche, con riferimento ai caratteri considerati.

Un fenomeno dicotomico può essere codificato in termini di variabile indicatrice che assume i valori:
1 = presenza del carattere (sì);
0 = assenza del carattere (no).

Con riferimento a due generiche unità statistiche ω_i e ω_j, i valori assunti dalle p variabili indicatrici, corrispondenti ai p caratteri dicotomici di partenza, possono essere classificati nella seguente tabella 2×2:

	ω_j		
ω_i	1	0	Totale
1	a	b	$a+b$
0	c	d	$c+d$
Totale	$a+c$	$b+d$	p

dove:

a = frequenza dei fenomeni contemporaneamente presenti nelle due unità (co-presenze o *potivive matches*)

d = frequenza dei fenomeni assenti in entrambe le unità (co-assenze o *negative matches*)

b, c = frequenze dei fenomeni presenti in un'unità ma non nell'altra.

Le frequenze b e c segnalano quindi gli aspetti di diversità tra le due unità statistiche considerate ed esse devono venire trattate allo stesso modo, per la condizione di simmetria. Le frequenze a e d indicano invece l'entità della rassomiglianza tra le due unità, ma il loro significato non è identico. Infatti la co-presenza di un carattere costituisce sempre un aspetto che concorre a definire una similarità, mentre la co-assenza di un fenomeno in alcuni casi può risultare di scarso o di nessun senso ai fini della valutazione della rassomiglianza fra le due unità.

Gli indici di similarità per fenomeni dicotomici vengono allora distinti a seconda che essi considerino o meno le co-assenze per la valutazione delle rassomiglianze.

Indici funzione delle co-presenze. Si definisce *indice di similarità di Russel e Rao* la quantità:

$$S_{RR} := \frac{a}{p} \tag{1.38}$$

che dunque risulta uguale al rapporto fra il numero di co-presenze ed il numero totale di caratteri considerati.

Si definisce *indice di similarità di Jaccard* la quantità:

$$S_J := \frac{a}{a+b+c} \tag{1.39}$$

che quindi risulta uguale al rapporto fra il numero di co-presenze ed il numero dei caratteri con esclusione di quelli che manifestano co-assenze. Si noti che questo indice risulta indeterminato se le due unità statistiche in esame presentano unicamente co-assenze.

Indici funzione delle co-presenze e delle co-assenze. Si definisce *indice di similarità di Sokal e Michener* la quantità:

$$S_{SM} := \frac{a+d}{p} \tag{1.40}$$

che è dato dal rapporto fra il numero di caratteri che risultano uguali nelle due unità ed il numero totale p di caratteri considerati.

Si può verificare che fra questi indici sussiste la seguente relazione:

$$S_{RR} \leq S_J \leq S_{SM} \, .$$

Si noti che gli indici suddetti differiscono unicamente per il diverso trattamento delle co-assenze, inotre sono tutti uguali se $d = 0$.

1.8.7 Indici di similarità per fenomeni con più di due modalità

Consideriamo ora il caso di fenomeni qualitativi con più di due modalità; parliamo di fenomeni politomici. Un importante tipo di applicazione concerne il calcolo della similarità quando si considerano le p domande di un questionario, con più modalità di risposta di tipo chiuso (alcune eventualmente di tipo dicotomico). Un indice di similarità riferito a due generici partecipanti all'inchiesta consente di valutare in quale misura differiscono tra loro due unità statistiche e quindi costituisce la premessa per l'individuazione di gruppi omogenei, o tipologie, di rispondenti.

Un criterio per calcolare la similarità tra due unità statistiche, con riferimento a p caratteri qualitativi X_i, ciascuno dei quali presenta $r_h > 2$ modalità ($h = 1, \ldots, p$) consiste nel tradurre i dati in termini di una matrice con codificazione disgiuntiva (vedi par. 1.4). Ciascun fenomeno viene presentato mediante r_h variabili indicatrici (che costituiscono la matrice di incidenza di dimensioni $n \times r_h$), ognuna delle quali assume valore 1 in corrispondenza delle unità che mostrano la modalità cui esse si riferiscono e valore 0 altrimenti. Per l'insieme delle n unità e dei p fenomeni si ottiene quindi una matrice booleana di dimensione $n \times R$, dove $R = \sum_h r_h$: ad ogni unità statistica corrisponde un vettore R-dimensionale, che contiene p valori uguali a 1 e $(R - p)$ uguali a 0.

Si considerino, ad esempio, tre fenomeni, di cui il primo può presentare 5 modalità, il secondo 3 ed il terzo 4; si assuma quindi che un'unità ω_i presenti la seconda modalità del primo carattere, la terza del secondo carattere e la prima del terzo carattere; l'untà ω_j presenta la seconda modalità del primo carattere, la prima del secondo fenomeno e la quarta del terzo. I vettori che codificano tali unità sono rispettivamente:

$$\begin{aligned} \mathbf{x}_i &= (0,1,0,0,0;0,0,1;1,0,0,0)' \\ \mathbf{x}_j &= (0,1,0,0,0;1,0,0;0,0,0,1)' \,. \end{aligned}$$

La similarità fra ω_i e ω_j viene colta ponendo a confronto i rispettivi vettori R-dimensionali le cui componenti possono essere classificate in una tabella 2×2 analogamente a quella riportata nel paragrafo 1.8.6 ma con il totale delle frequenze pari a R:

	ω_j		
ω_i	1	0	Totale
1	α	β	$\alpha + \beta$
0	γ	δ	$\gamma + \delta$
Totale	$\alpha + \gamma$	$\beta + \delta$	R

Si noti che, in questo caso, le co-assenze non forniscono alcuna informazione sulla similarità esistente, in quanto segnalano unicamente che nelle due unità non compare la modalità considerata, che è una delle $r_h > 2$ possibili. Pertanto, gli indici di similarità ora logicamente utilizzabili sono solo quelli che esprimono la rassomiglianza come funzione esclusivamente delle co-presenze.

Un primo indice impiegabile, solo formalente analogo a quello di Jaccard, è il seguente:

$$_A S_J := \frac{\alpha}{\alpha + \beta + \gamma}$$

in cui le coppie 0-0 sono completamente ignorate. Si osservi però che, per ogni fenomeno politomico con codificazione disgiuntiva, la mancata coincidenza della modalità nelle

due unità statistiche considerate comporta sempre le due coppie 1-0 e 0-1 nelle variabili indicatrici, corrispondenti rispettivamente alle modalità assunte dal primo elemento e dal secondo elemento. Questo genera una duplicazione delle frequenze nella tabella 2×2. Nell'esempio sopra riportato, le unità ω_i e ω_j mostrano il primo fenomeno una co-presenza della seconda modalità e questo comporta un incremento unitario della frequenza α, mentre il secondo fenomeno presenta modalità differenti e ciò comporta un incremento unitario sia di β (coppia 1-0) sia di γ (coppia 0-1); la stessa cosa di verifica per il terzo fenomeno.

In questa situazione, sembra quindi ragionevole attribuire per 1/2 alle frequenze β e γ, definendo il seguente indice di similarità:

$$ {}_BS_J := \frac{\alpha}{\alpha + 0.5(\beta + \gamma)} $$

che ha un'interpretazione immediata: è il rapporto fra il numero di fenomeni che presentano modalità coincidenti nelle due unità statistiche rispetto al numero totale dei fenomeni considerati.

1.9 Matrici di preferenza

Consideriamo il caso in cui ad ogni unità statistica si richiede di classificare in ordine di preferenza p elementi di un insieme, quali ad esempio le p diverse marche di un prodotto, p spettacoli televisivi, p diversi sport, etc. In tal caso x_{hi} costituisce il grado di preferenza espresso dalla h-esima unità statistica con riferimento all'i-esima variabile. La matrice $\underset{\sim}{\mathbf{X}}$ che si ottiene viene chiamata *matrice di preferenze*. Le variabili $X_1, \ldots, X_p$ sono variabili categoriali ordinali; lo studio della dipendenza fra due unità statistiche di tale matrice viene effettuato mediante *indici di cograduazione*.

1.9.1 La cograduazione (caso senza unità alla pari)

Consideriamo prima il caso in cui in ognuna delle due graduatorie non vi siano unità alla pari, cioè non vi siano unità che occupino lo stesso posto in una graduatoria (per esempio due distinte unità non possono occupare il terzo posto). La tabella seguente riporta le valutazione ottenute da alcuni insegnamenti del Corso di Laurea in Scienze Statistiche ed Attuariali nell'anno accademico 2000-2001 per le variabili *organizzazione del corso* X ed *efficacia lezione* Y:

codice corso	*organizzazione corso*	*efficacia lezione*
314	7.9	7.2
315	6.6	6.7
317	7.6	6.3
319	8.2	8.1
333	7.4	5.4
337	7.2	7.4
349	8.7	7.6

Ci poniamo il problema di valutare la concordanza fra i giudizi espressi per le due variabili in esame. La matrice sopra riportata è un semplice esempio di *matrice di punteggi* dove il generico elemento x_{hi} rappresenta il punteggio attribuito dalla h-esima unità statistica per la i-esima variabile.

Siano X e Y due variabili aventi ciascuna n modalità. Ci si pone il problema di misurare la relazione fra X e Y analizzando i rispettivi *posti d'ordine* o *gradi* o *ranghi* (*rank*). Si assegna pertanto grado "1" alla modalità inferiore, grado "2" a quella immediatamente successiva, e così via grado n alla modalità maggiore. La metodologia statistica che studia le relazioni tra i posti d'ordine delle modalità di caratteri ordinali o quantitativi si chiama *cograduazione* (*rank correlation*).

Assumiamo preliminarmente che le n modalità di X e di Y siano distinte, così che non vi siano *ex-aequo*. Consideriamo una funzione $g(\cdot)$ che associa a ciascuna modalità il suo rango, cioè $g : \mathcal{X} \to \{1, 2, \ldots, n\}$. Indichiamo con $x_{(1)}$ la più piccola modalità assunta da X in Ω, con $x_{(2)}$ quella immediatamente successiva e così via:

$$x_{(1)} < x_{(2)} < \cdots < x_{(n)}$$

e quindi scriveremo:

$$g(x_i) = j \qquad i = 1, \ldots, n \ .$$

se l'osservazione x_i ha posto j nella graduatoria in esame; in altre parole $g(x_i)$ fornisce il posto nella graduatoria delle modalità del valore osservato x_i. Con riferimento all'esempio precedente, il corso avente codice 314 ha ottenuto una valutazione pari a 7.9 per l'organizzazione del corso (corrispondente al quinto posto in graduatoria) ed una valutazione pari a 7.2 per l'efficacia della lezione (pari al quarto posto in graduatoria):

		organizzazione corso X		*efficacia lezione* Y	
i	*codice corso*	punteggio	graduatoria	punteggio	graduatoria
1	314	7.9	5	7.2	4
2	315	6.6	1	6.7	3
3	317	7.6	4	6.3	2
4	319	8.2	6	8.1	7
5	333	7.4	3	5.4	1
6	337	7.2	2	7.4	5
7	349	8.7	7	7.6	6

Le quantità $g(x_i)$ vengono chiamati "ranghi" o "gradi". Se sostituiamo alle modalità x_i e y_i i rispettivi ranghi (in ordine crescente), lo studio della relazione fra X e Y può essere ricondotto allo studio della relazione fra i vettori:

$$\begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_n) \end{pmatrix} \qquad \text{e} \qquad \begin{pmatrix} g(y_1) \\ g(y_2) \\ \vdots \\ g(y_n) \end{pmatrix}$$

ciascuno dei quali contiene i numeri da 1 a n in un ordine che dipende dalle modalità dei fenomeni. Introduciamo ora i due casi limite.

Definizione 1.8 Diremo che fra i due fenomeni X e Y vi è *perfetta cograduazione* se risulta:

$$g(x_i) = g(y_i) \qquad i = 1, \ldots, n \ .$$

cioè quando le unità statistiche, con riferimento ai due caratteri X e Y, presentano lo stesso ordinamento nelle due graduatorie (una certa unità statistica assume il primo posto nella graduatoria di X ed il primo posto nella graduatoria di Y; un'altra unità statistica assume il secondo posto nella graduatoria sia di X che di Y, e così via).

Diremo che fra i due fenomeni X e Y vi è *perfetta contrograduazione* se risulta

$$g(x_i) = n + 1 - g(y_i) \qquad i = 1, \ldots, n \ .$$

cioè quando le unità statistiche, con riferimento ai due caratteri X e Y, presentano ordinamento opposto nelle due graduatorie (una certa unità statistica assume il primo posto nella graduatoria di X e l'ultimo posto nella graduatoria di Y; un'altra unità statistica assume il secondo posto nella graduatoria di X ed il penultimo in quella di Y, e così via)).

La relazione fra i due fenomeni può pertanto esere misurata facendo riferimento alle rispettive graduatorie, cioè costruendo indici che basati sulle graduatorie.

Nota 1.9 Con riferimento ad una matrice di dati di dimensioni $n \times p$, se essa contiene fenomeni sia quantitativi che ordinali, la trasformazione degli stessi in termini delle rispettive graduatorie rappresenta inoltre un criterio per il trattamento simultaneo dei caratteri. Ovviamente per i fenomeni quantitativi ciò comporta una perdita di informazione dovuta al passaggio da una scala di misurazione quantitativa ad una scala di misurazione ordinale.

Si considerino due coppie di gradi riferite alle generiche unità i e j:

$$\{g(x_i), g(y_i)\} \qquad \text{e} \qquad \{g(x_j), g(y_j)\} \ .$$

Tali coppie si dicono *concordanti* se l'unità che presenta grado maggiore su X presenta grado maggiore su Y, cioè se risulta:

$$g(x_i) < g(y_i) \quad \text{e} \quad g(x_j) < g(y_j) \ ,$$

oppure

$$g(x_i) > g(y_i) \quad \text{e} \quad g(x_j) > g(y_j) \ .$$

Due coppie di gradi si dicono *discordanti* se l'unità che presenta grado maggiore su X presenta grado minore su Y, cioè se risulta:

$$g(x_i) < g(y_i) \quad \text{e} \quad g(x_j) > g(y_j) \ ,$$

oppure

$$g(x_i) > g(y_i) \quad \text{e} \quad g(x_j) < g(y_j) \ .$$

Ad esempio, con riferimento ai dati presentati all'inizio, le unità che si riferiscono ai corsi con codice 315 (ω_2) e 337 (ω_6) sono concordanti, in quanto $g(x_2) = 1 < g(y_2) = 3$ e $g(x_6) = 2 < g(y_6) = 5$; al contrario i corsi con codice 315 (ω_2) e 317 (ω_3) sono discordanti in quanto $g(x_2) = 1 < g(y_2) = 3$ e $g(x_3) = 4 > g(y_3) = 2$.

1.9.2 L'indice di Kendall (caso senza unità alla pari)

Si considerino le graduatorie di n unità statistiche, riferite alle modalità di due caratteri ordinali X e Y. Senza perdere di generalità, possiamo assumere che le unità statistiche siano disposte seguendo l'ordine naturale di X. Ne segue che la misura della concordanza fra X e Y può essere colta in base al modo secondo il quale si succedono i corrispondenti gradi del carattere Y:

- se esiste una perfetta cograduazione, i gradi di Y costituiscono anch'essi una successione ordinata da 1 a n;
- se esiste una perfetta contrograduazione, allora i gradi di Y si susseguono nell'ordine opposto, cioè da n a 1.

In generale una misura di cograduazione può essere costruita come segue. Considerate due doppie

Una misura di cograduazione può allora essere definita in base al numero di inversioni della graduatoria di Y rispetto a quella di X.

Definizione 1.10 Si dice che vi è un'inversione quando, disponendo le unità in modo che i gradi di X seguano l'ordine naturale da 1 a n e confrontando il grado d'un termine generico $g(y_i)$ con il grado di un altro termine che lo segue nella successione, risulta:

$$g(y_i) > g(y_j) \qquad j > i; \quad i = 1, \ldots, n$$

Riprendiamo l'esempio iniziale, dove X="organizzazione del corso" e Y="efficacia lezione"; i dati sono riportati nella seguente tabella:

	organizzazione corso		*efficacia lezione*	
codice corso	punteggio	graduatoria	punteggio	graduatoria
315	6.6	1	6.7	3
337	7.2	2	7.4	5
333	7.4	3	5.4	1
317	7.6	4	6.3	2
314	7.9	5	7.2	4
319	8.2	6	8.1	7
349	8.7	7	7.6	6

Al fine di misurare la concordanza fra le due graduatorie, consideriamo le seguenti quantità con riferimento a due qualunque unità statistiche ω_i e ω_j ($i, j = 1, \ldots, n$, $i \neq j$):

$$d_{ij} = \begin{cases} +1 & \text{se } g(x_i) < g(x_j) \\ -1 & \text{se } g(x_i) > g(x_j) \end{cases} \qquad \text{e} \qquad \delta_{ij} = \begin{cases} +1 & \text{se } g(y_i) < g(y_j) \\ -1 & \text{se } g(y_i) > g(y_j) \end{cases} \tag{1.41}$$

e consideriamo l'insieme di tutti i possibili confronti fra coppie di unità statistiche. In altre parole, confrontiamo il rango di ciascuna unità statistica ω_i con quello di tutte le altre

unità statistiche ω_j ($j \neq i$). Il numero totale di confronti è $n(n-1)/2$. Successivamente si costruisce il seguente indice, noto come τ di Kendall:

$$\tau := \frac{\sum_{i<j} d_{ij}\delta_{ij}}{n(n-1)/2} \,. \tag{1.42}$$

Ovviamente si tratta di un indice normalizzato che assume valore fra -1 (nel caso in cui $d_{ij}\delta_{ij} = -1$ per tutte le coppie $i < j$) e $+1$ (nel caso in cui $d_{ij}\delta_{ij} = +1$ per tutte le coppie $i < j$). In particolare risulta:

- $\tau = -1$ quando vi è perfetta contrograduazione fra i due fenomeni in esame;
- $\tau = 0$ quando i due fenomeni in esame non presentano né concordanza né discordanza;
- $\tau = +1$ quando vi è perfetta graduazione fra i due fenomeni in esame.

Inoltre l'indice di cograduazione di Kendall è simmetrico: $\tau(X, Y) = \tau(Y, X)$.

Proprietà dell'indice τ di Kendall. L'indice τ di Kendall è invariante per trasformazioni monotone crescenti delle graduatorie dei due caratteri. L'assegnazione dei numeri da 1 a n alle modalità ordinali di un fenomeno è convenzionale. In alternativa, si potrebbero assegnare all'uno o all'altro fenomeno gradi crescenti (o ad entrambi), ad esempio secondo una funzione quadratica: $g(x_{(i)} = i^2$, oppure con progressione geometrica: $g(x_{(i)} = a^i$ (per $a > 1$ e $i = 1, \ldots, n$). L'indice τ non risulta affatto influenzato da tali assegnazioni in quanto, in base alla (1.41), il confronto nella definizione (1.42) si basa unicamente sulle relazioni di maggiore o minore fra le modalità, che rimangono immutate a seguito di una qualunque trasformazione monotona crescente dei valori numerici assunti come gradi.

Con riferimento al caso in esame, essendo $n = 7$ si hanno 21 confronti, che vengono riportati nella seguente tabella:

ω_i	ω_j	d_{ij}	δ_{ij}	$d_{ij}\delta_{ij}$
1	2	-1	-1	1
1	3	-1	-1	1
1	4	1	1	1
1	5	-1	-1	1
1	6	-1	1	-1
1	7	1	1	1
2	3	1	-1	-1
2	4	1	1	1
2	5	1	-1	-1
2	6	1	1	1
2	7	1	1	1

ω_i	ω_j	d_{ij}	δ_{ij}	$d_{ij}\delta_{ij}$
3	4	1	1	1
3	5	-1	-1	1
3	6	-1	1	-1
3	7	1	1	1
4	5	-1	-1	1
4	6	-1	-1	1
4	7	1	-1	-1
5	6	-1	1	-1
5	7	1	1	1
6	7	1	1	1

e pertanto segue:

$$\tau = \frac{9}{21} = 0.4286 \,.$$

1.9.3 Indici di Spearman (caso senza unità alla pari)

Un altro indice di cograduazione è il seguente. Con riferimento alla (1.41) poniamo:

$$d_{ij} := g(x_j) - g(x_i) \qquad \delta_{ij} := g(y_j) - g(y_i) \; . \tag{1.43}$$

L'*indice di cograduazione di Spearman*, usualmente denotato con la lettera ρ, è il seguente:

$$\rho := \frac{\sum_i \sum_j [g(x_j) - g(x_i)][g(y_j) - g(y_i)]}{\sqrt{\sum_i \sum_j [g(x_j) - g(x_i)]^2 \sum_i \sum_j [g(y_j) - g(y_i)]}} \; . \tag{1.44}$$

In particolare risulta:

- $\rho = -1$ quando vi è perfetta contrograduazione fra i due fenomeni in esame;
- $\rho = 0$ quando i due fenomeni in esame non presentano né concordanza né discordanza;
- $\rho = +1$ quando vi è perfetta graduazione fra i due fenomeni in esame.

Inotre, ovviamente, in base alla (1.44, l'indice di cograduazione di Spearman è simmetrico, cioè $\tau(X, Y) = \tau(Y, X)$.

Ai fini pratici risulta più agevole la seguente forma di ρ (vedi [Let83] per dettagli):

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} [g(x_i) - g(y_i)]^2}{n(n^2 - 1)} \tag{1.45}$$

che risulta più agevole nei casi pratici.

Nota 1.11 Si può dimostrare che il coefficiente di cograduazione di Spearman fra due graduatorie (senza unità alla pari) è numericamente uguale al coefficiente di correlazione lineare fra due distribuzioni aventi come modalità corrispondenti i posti occupati nelle due graduatorie da una stessa unità (vedi [Let83] per approfondimenti) in quanto si dimostra che risulta:

$$\rho = \frac{\sum_i [g(x_i) - \frac{n+1}{2}][g(y_i) - \frac{n+1}{2}]}{\sqrt{\sum_i [g(x_i) - \frac{n+1}{2}]^2 \sum_i \sum_j [g(y_i) - \frac{n+1}{2}]}} \; . \tag{1.46}$$

e la quantità $\frac{n+1}{2}$ si può interpretare come la media aritmetica di $g(x_1), \ldots, g(x_n)$ e di $g(y_1), \ldots, g(y_n)$ che sono permutazioni dei numeri $1, \ldots, n$.

Proprietà dell'indice ρ di Spearman. In base alla definizione (1.44), si vede immediatamente che l'indice di cograduazione ρ di Spearman è invariante per trasformazioni monotone crescenti delle modalità di uno o di entrambi i fenomeni. Infatti, qualunque modificazione delle modalità originarie di un carattere che non alteri i posti d'origine, non altera la graduatoria che conduce al calcolo dell'indice stesso. Si noti che questa proprietà di invarianza è più debole di quella dell'indice τ in quanto non è lecito trasformare monotonicamente i gradi, bensì solo le modalità originarie, sulle quali si determinano sempre le graduatorie, del tipo consueto da 1 a n.

Inoltre, in base alla (1.46), poichè ρ è uguale al coefficiente di correlazione calcolato sulle distribuzioni dei posti d'ordine, esso è invariante solo per trasformazioni lineari crescenti delle graduatorie.

Confronto fra l'indice τ di Kendall e l'indice ρ di Spearman. Nell'indice ρ l'invarianza è raggiunta imponendo che la misurazione assegni valore 1 alla modalità minore di tutte le altre, valore 2 a quella immediatamente successiva, etc. La scala di misurazione è rigidamente imposta ed è possibile trasformarla solo linearmente: non si può più parlare allora, legittimamente, di scala ordinale. L'assunzione dei primi n numeri naturali come misure delle n modalità di ciascun vettore implica dunque l'ipotesi addizionale, usualmente non enunciata esplicitamente, di *uguale distanza fra i gradi* che traducono numericamente l'ordinamento delle modalità. Nel caso di fenomeni qualitativi con modalità ordinabili, ciò è in contrasto con la scala di misurazione, che ammette soltanto le relazioni di "uguale", "maggione" e "minore". Nel caso di fenomeni (quantitativi) misurabili su scala d'intervalli o su scala di rapporti, la trasformazione dei valori originari in gradi secondo la regola sopra descritta comporta l'arbitraria riduzione a 1 di tutte le differenze effettive fra valori consecutivi. Il ricorso all'indice ρ non significa dunque, in realtà, valutare semplicemente la concordanza, tra gli ordinamenti delle modalità dei due vettori componenti la serie doppia, come spesso si dice, bensì quantificare la relazione fra misurazioni (i gradi) i cui valori costituiscono una progressione aritmetica di ragione 1 e con primo termine anch'esso uguale a 1.

Nell'indice τ di Kendall, invece, l'impiego delle graduatorie ha l'unico scopo di semplificare i calcoli.

La relazione fra il valore numerico di ρ e quello di τ calcolati sulle medesime graduatorie non è affatto semplice, se si escludono i casi banali di perfetta cograduazione e perfetta contrograduazione in cui entrambi assumono valore rispettivamente uguale a $+1$ e -1. Con larga approssimazione, si può dire che il rapporto fra ρ e τ si avvicina a 3/2 per n grande e per valori degli indici non troppo prossimi agli estremi unitari

Relazione fra indici di cograduazione e correlazione. Gli indici di cograduazione, che sono applicabili a caratteri unicamente ordinali, sono calcolabili anche in presenza di fenomeni quantitativi, per i quali è sempre determinabile la corrispondente graduatoria. In quest'ultima circostanza, per lo studio delle relazioni fra due variabili, gli indici di cograduazione rappresentano un criterio alternativo rispetto al consueto coefficiente di correlazione lineare r introdotto in (1.12). Risulta importante mettere quindi in luce le differenze fra i vari indici.

In primo luogo, si può dimostrare che gli indici τ, ρ e r sono tutti riconducibili ad una stessa formula generale, vedi [Let83]:

$$\gamma := \frac{\sum_i \sum_j \alpha_{ij}\beta_{ij}}{\sqrt{\sum_i \sum_j \alpha_{ij}^2 \sum_i \sum_j \beta_{ij}^2}} \qquad \text{con } \alpha_i = \beta_i = 0 \tag{1.47}$$

dove α_{ij} rappresenta un punteggio di un certo tipo, assegnato alla coppia di unità ω_i e ω_j con riferimento al carattere X e β_{ij} ha un significato analogo, con riferimento al carattere Y. Si assume che $\alpha_{ij} = -\alpha_{ji}$ e $\beta_{ij} = -\beta_{ji}$.

1. L'indice τ di Kendall si ottiene dalla (1.47) assegnando i punteggi nella maniera seguente:

$$\begin{aligned} a_{ij} &= +1 \qquad \text{se } g(x_i) < g(x_j) \\ a_{ij} &= -1 \qquad \text{se } g(x_i) > g(x_j)\,, \end{aligned}$$

   e similmente per b_{ij} con riferimento ai gradi di Y.