

multidimensional scaling metrico

Si è soliti parlare di multidimensional scaling metrico quando i valori di prossimità in input sono di tipo metrico, ossia sono espressi secondo scale di misura ad intervallo o a rapporti.

In questo caso l'obiettivo dell'analisi è di determinare una configurazione degli n elementi in uno spazio geometrico di dimensione ridotta, in modo tale che le distanze tra i punti che li rappresentano siano il più possibile vicine alle prossimità originarie.

In generale all'aumentare della non somiglianza, deve crescere anche la distanza tra i punti nello spazio geometrico.

Si possono utilizzare diverse misure di distanza tra i punti; ma quella maggiormente usata è la distanza euclidea.

Nella pratica delle ricerche di mercato matrici di dati di prossimità di questo tipo sono raramente disponibili.

multidimensional scaling metrico: scelta del numero di dimensioni

Il modello cosiddetto metrico, utilizzabili per dati di prossimità su scala metrica ed espressi in forma aggregata, ha lo scopo di individuare una configurazione degli n prodotti o marche in esame in uno spazio geometrico di dimensioni p .

Ai fini dell'interpretazione di una soluzione di MDS è cruciale proprio la scelta del numero di dimensioni della configurazione.

Un numero limitato facilita la rappresentazione grafica facilitando di conseguenza l'interpretazione.

Ciò non garantisce però che la soluzione sia anche statisticamente soddisfacente.

Una regola di comportamento pratico suggerisce di tener conto del numero di prodotti o marche sottoposti all'analisi.

Per ottenere una configurazione interpretabile è necessario che le coppie distinte di elementi presi in esame siano almeno in numero doppio rispetto alle coordinate della mappa da stimare.

determinazione delle coordinate della mappa

Dati quindi N elementi, di cui si conosce la matrice delle distanze $D(N,N)$, la tecnica MDS ci permette di definire una matrice di coordinate $X(N,k)$, con $k \ll N$) in uno spazio metrico euclideo adeguato a rappresentare gli N elementi considerati. Tali coordinate non sono quelle dei punti realmente associati alla matrice D originaria e che in seguito saranno indicati con Z . Ma esse sono relative ad un insieme di punti tali che la matrice delle loro distanze D sia il più possibile simile alla matrice D originaria. Per dare una rappresentazione centrata delle coordinate dei punti si introduce la matrice B così definita:

$$b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})' (\mathbf{z}_s - \bar{\mathbf{z}})$$

b_{rs} rappresenta le coordinate del r -esimo ed s -esimo punto, individuati come elementi della matrice Z .

Inoltre si osservi che, diagonalizzando la matrice B si ottiene:

$$\mathbf{B} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}'$$

La simmetria della matrice B garantisce l'esistenza di una base ortonormale di autovettori Γ per B e permette di definire la matrice Λ come una matrice diagonale in cui sono presenti gli autovalori della matrice B .

Quindi è possibile arrivare alla formulazione X studiando gli autovalori e gli autovettori della matrice B , ovvero:

$$X = \begin{matrix} & \lambda_1 & \lambda_2 & \dots & \lambda_q \\ \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1q} \\ x_{21} & x_{22} & \dots & x_{2q} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nq} \end{bmatrix} & P_1 & P_2 & \vdots & P_N \end{matrix}$$

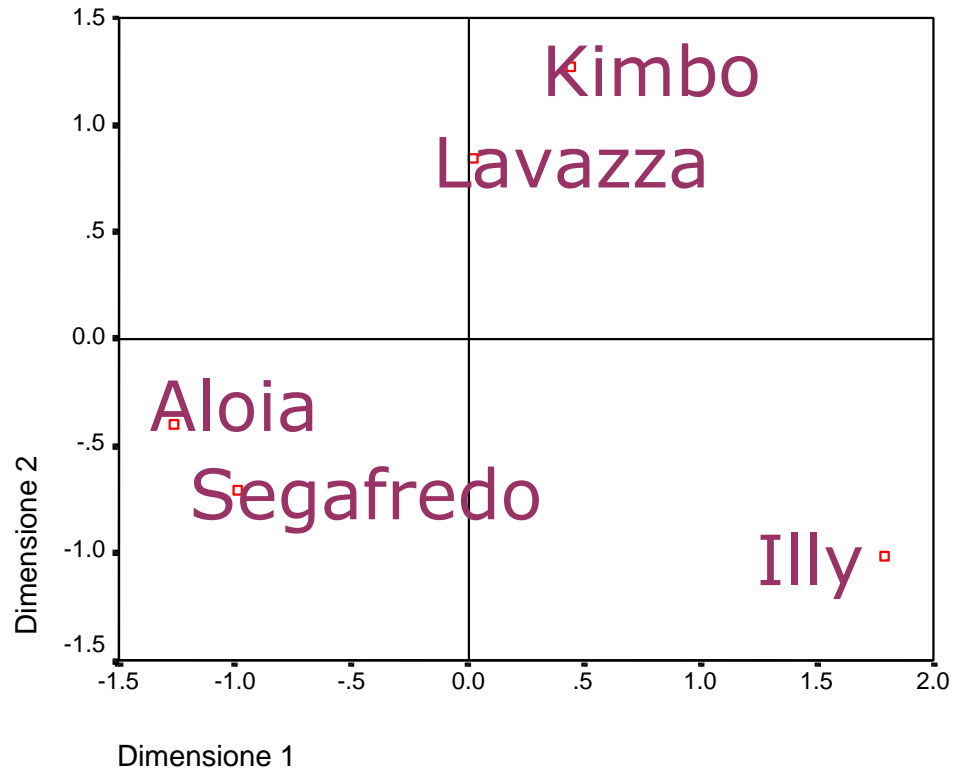
dove le colonne rappresentano gli autovettori calcolati mentre la riga i -esima costituisce le coordinate dell' i -esimo elemento che si voleva rappresentare.

Esempio “la caffè do brasil”

La Café do Brasil s.p.a., azienda partenopea fondata dalla famiglia Rubino nel 1950 leader nella produzione di caffè, intende conoscere il modo in cui il proprio prodotto si posiziona nella mente dei consumatori rispetto alle altre marche di caffè presenti sul mercato. Le marche dei principali concorrenti appartenenti alla stessa classe di prodotti sono:

- Kimbo
- Lavazza,
- Illy,
- Segafredo,
- Aloia

	Raccolta punti	Prezzo	Qualità	Reperibil ità	Varietà	Publicità	Numero Verde	Sponsor
Kimbo	no	1,8	Medio/alta	Alta	Medio/alta	si	no	no
Lavazza	si	2,32	Medio/alta	Alta	Alta	si	no	si
Segafredo	no	1,55	Medio/bassa	Media	Media	si	si	si
Illy	no	5	Alta	Bassa	Bassa	si	si	no
Aloia	no	1,03	Bassa	Bassa	Bassa	no	no	no



1	KIMBO	.4419	1.2726
2	LAVAZZA	.0261	.8410
3	SEGAFFRED	-.9928	-.7007
4	ILLY	1.7908	-1.0168
5	ALOIA	-1.2661	-.3962

Sulla prima dimensione si ha la contrapposizione tra Illy e le altre marche, in particolare Aloia e Segafredo.

Gli attributi che differenziano la Illy dalla Aloia e dalla Segafredo sono il prezzo e la qualità.

La prima dimensione sintetizza il rapporto prezzo-qualità

Seconda dimensione: contrapposizione tra Lavazza e Kimbo, da una parte, e Illy dall'altra.

Il contrasto si basa su: prezzo, reperibilità, varietà dei prodotti, ricorso a pubblicità televisiva e esistenza di un servizio consumatori (numero verde) e ancora una volta qualità.

Significato alla seconda dimensione: si può parlare di variabile di status.

multidimensional scaling non metrico

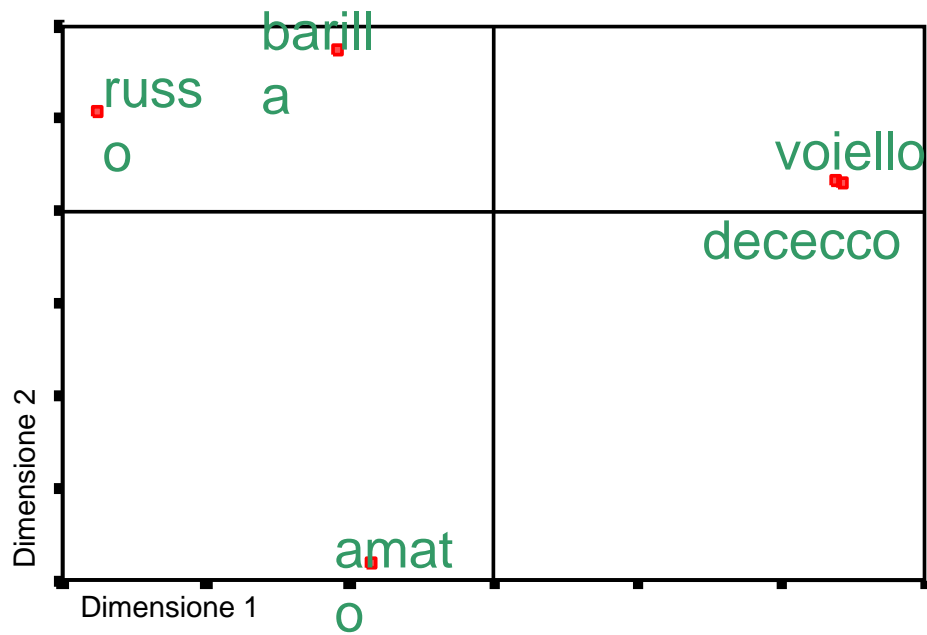
Se i valori di prossimità in input sono espressi su una scala di tipo ordinale si applicano metodi di Multidimensional scaling non metrico. In questo caso l'obiettivo dell'analisi è di pervenire ad una configurazione degli n elementi in uno spazio geometrico di dimensione ridotta tale per cui l'ordinamento generato dalle distanze tra gli n punti che li rappresentano sia il più possibile analogo a quello dato dalle prossimità originarie.

Con l'MDS non metrico non è possibile definire direttamente una soluzione analitica, ma viene utilizzata una procedura iterativa, che, partendo da una configurazione iniziale di distanze, scelta a caso, consente di giungere ad una matrice finale di distanze e di coordinate.

L'esempio riguarda un'indagine sul posizionamento di 5 marche di pasta alimentare. Il giudizio di prossimità dei consumatori poteva essere distorto dal differente e vasto assortimento esistente nell'ambito della stessa marca. Per tale motivo, il confronto è stato limitato ad un singolo prodotto: gli spaghetti, che rappresenta un pò il biglietto da vista delle marche considerate almeno per il mercato territorio d'indagine, che è quello napoletano. Le marche prese in considerazione sono: Voiello, Barilla, De Cecco, Russo e Amato.

Caratteristiche	Marche				
	Voiello	Barilla	De Cecco	Russo	Amato
Prezzo (Kg)	1,40	1,11	1,34	0,93	1,03
Qualità	Alta	Alta	Alta	Bassa	Media
Pubblicità Televisiva	Nazionale	Nazionale	Nazionale	Locale	Assente
Reperibilità	Media	Alta	Media	Alta	Alta
Cottura	8'-12'	10'-15'	10'-15'	7'-13'	8'-13'
Tenuta Cottura	Media	Alta	Alta	Bassa	Bassa

Stress = ,00307



Matrice delle coordinate delle marche

1	VOIELLO	1,2022	,1847
2	BARILLA	-,5583	,9131
3	DECECCO	1,1772	,1993
4	RUSSO	-1,3868	,5755
5	AMATO	-,4342	-1,8726

Matrice delle coordinate delle marche

1	VOIELLO	1,2022	,1847
2	BARILLA	-,5583	,9131
3	DECECCO	1,1772	,1993
4	RUSSO	-1,3868	,5755
5	AMATO	-,4342	-1,8726

La prima dimensione sintetizza il prezzo con la reperibilità
La seconda sintetizza l'immagine di marca che scaturisce dalla comunicazione pubblicitaria.

Le fasi del multidimensional scaling non metrico

1. Dalla matrice delle dissimilarità, si crea una configurazione iniziale;
2. si calcolano le distanze tra tutte le possibili coppie di elementi della precedente configurazione;
3. si confronta l'ordinamento di tali distanze con quello degli indici di dissimilarità per vedere se è soddisfatta l'ipotesi di monotonicità;
4. se la relazione di monotonicità non è rispettata, si calcolano le disparità (o pseudo distanze), con diversi metodi, ad esempio con la regressione monotonica, che consiste nel trasformare le distanze in modo che sia rispettato l'ordinamento delle dissimilarità.

In pratica, ordinate le distanze secondo il posto occupato dai rispettivi indici di dissimilarità, se nella successione delle distanze alcuni valori consecutivi si discostano dalla monotonicità, ad ognuno di essi si sostituisce la loro media.

La nuova successione, se è monotona, è quella delle disparità, le quali non sono distanze, se non è monotona, si ripete il procedimento finché non viene soddisfatta la condizione di monotonicità.

Nel multidimensional scaling non metrico la qualità dell'adattamento ai dati viene misurato trami una misura chiamata stress:

$$S = \left[\frac{\sum_{i < j} \sum_j (d_{ij} - f_{ij})^2}{\sum_{i < j} \sum_j (d_{ij})^2} \right]^{1/2}$$

La scala di Likert 1

La scala di **Likert**, sviluppata nell'ambito della psicometria, si usa frequentemente nei questionari quando si vogliono risposte circa il livello di accordo nei confronti di una affermazione.

Una tipica domanda Likert, su una scala a 5, ha questa forma:

- Decisamente contrario
- Contrario
- Non contrario né favorevole
- Favorevole
- Decisamente favorevole

Di solito si usano scale a 4, 5, 7, oppure a 10

Occorre porre attenzione alla natura della risposta che non è una reale misura, da interpretare in senso metrico. Talvolta è una forzatura persino affidarsi alla natura ordinale delle misurazioni su una scala Likert.

Trattandosi di atteggiamenti, oppure di opinioni, possono presentarsi fenomeni per i quali i diversi elementi A , B , C possono presentare relazioni di circolarità:

$$A > B$$

$$B > C$$

$$C > A$$

violando l'assioma della **transitività** di una scala ordinale

La scala di Likert 2

Le principali fonti di distorsione a questo tipo di domande sono di solito associate ad alcuni atteggiamenti psicologici dei rispondenti, quali la tendenza:

1. ad evitare risposte estreme;
2. a collocarsi al centro (se è possibile)
3. ad essere d'accordo (distorsione di accondiscendenza)
4. a mettersi in buona luce con l'intervistatore (distorsione di desiderabilità sociale)

Progettare una scala con un buon equilibrio nelle risposte può ridurre la distorsione "da accondiscendenza", poiché l'accondiscendenza verso il versante positivo verrebbe ad essere bilanciata da accondiscendenza sul versante negativo.

La tendenza a posizionarsi al centro e la distorsione da "desiderabilità sociale" sono più difficile da controllare

unfolding

Il modello di analisi dei dati di preferenza conosciuto come *unfolding* ipotizza che i consumatori, al momento in cui formulano giudizi di preferenza, siano caratterizzati da un punto di **riferimento ideale**

Si stimano, quindi, anche le coordinate del punto ideale individuale:

- ogni coordinata designa il livello della caratteristica/attributo associabile alla dimensione di rappresentazione ideale per il soggetto
- quanto più un prodotto o una marca sono vicini al punto di riferimento ideale tanto più sono preferiti

Per valutare le coordinate dei punti ideali si possono utilizzare due distinti approcci, l'approccio implicito ed esplicito. Nel primo caso la stima delle coordinate del punto ideale avviene a partire dai soli giudizi di preferenza espressi sui prodotti a confronto

Nel secondo caso la stima è effettuata a partire da risposte dirette degli intervistati su questo argomento, il che avviene includendo un prodotto ideale ipotetico tra quelli giudicati