

Lezione : Approfondimenti ed esercizi

Docente: Aldo Solari

1 L'analisi dei gruppi

Example 1.1. Distanza tra gruppi: legame completo.Passo ①: Inizializzare $k = n$ e $D = D_{k \times k} = D_{n \times n}$

$$D_{5 \times 5} = \{d_{IL}\} =$$

$I \backslash L$	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

ITERAZIONE 1

② $\min_{I \neq L}(d_{IL}) = d_{53} = 2$

- Le due unità (cluster) 3 e 5 vengono fuse nel cluster (35)

③ Aggiorno le distanze tra il nuovo cluster (35) e i rimanenti

- $d_{(35)1} = \max\{d_{31}, d_{51}\} = \max\{3, 11\} = 11$
- $d_{(35)2} = \max\{d_{32}, d_{52}\} = \max\{7, 10\} = 10$
- $d_{(35)4} = \max\{d_{34}, d_{54}\} = \max\{9, 8\} = 9$

dove il legame completo $d_{(IL)J} = \max\{d_{IJ}, d_{LJ}\}$

$$D_{4 \times 4} = \{d_{IL}\} =$$

$I \backslash L$	(35)	1	2	4
(35)	0			
1	11	0		
2	10	9	0	
4	9	6	5	0

ITERAZIONE 2

② $\min_{I \neq L}(d_{IL}) = d_{42} = 5$

- I due cluster 2 e 4 vengono fusi nel cluster (24)

③ Aggiorno le distanze tra il nuovo cluster (24) e i rimanenti

- $d_{(24)(35)} = \max\{d_{2(35)}, d_{4(35)}\} = \max\{10, 9\} = 10$
- $d_{(24)1} = \max\{d_{21}, d_{41}\} = \max\{9, 6\} = 9$

$$D_{3 \times 3} = \{d_{IL}\} = \begin{array}{c|ccc} I \backslash L & (35) & (24) & 1 \\ \hline (35) & 0 & & \\ (24) & 10 & 0 & \\ 1 & 11 & \mathbf{9} & 0 \end{array}$$

ITERAZIONE 3

② $\min_{I \neq L}(d_{IL}) = d_{1(24)} = 9$

- I due cluster 1 e (24) vengono fusi nel cluster (124)

③ Aggiorno le distanze tra il nuovo cluster (124) e il rimanente

- $d_{(124)(35)} = \max\{d_{1(35)}, d_{(24)(35)}\} = \max\{11, 10\} = 11$

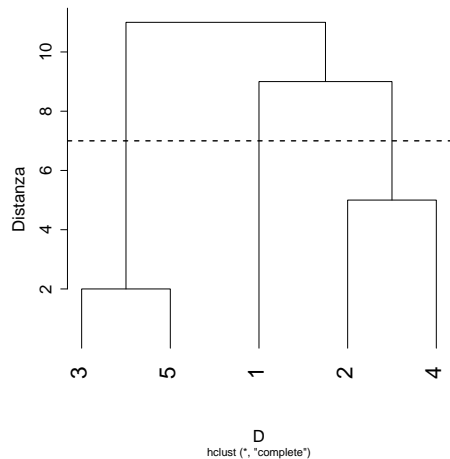
$$D_{2 \times 2} = \{d_{IL}\} = \begin{array}{c|cc} I \backslash L & (35) & (124) \\ \hline (35) & 0 & \\ (124) & \mathbf{11} & 0 \end{array}$$

ITERAZIONE 4

② $\min_{I \neq L}(d_{IL}) = d_{(35)(124)} = 11$

- I due cluster (35) e (124) vengono fusi nel cluster (12345)

③ STOP



2 Analisi fattoriale

Example 2.1. Analisi fattoriale

La seguente tabella riporta le stime di un modello fattoriale ottenute, previa standardizzazione dei dati, da misurazioni di alcune sostanze in bacini idrici.

	Factor1	Factor2
Calcio	0.453	
Magnesio	0.137	0.722
Sodio	0.942	0.264
Cloruri	0.827	0.222
Nitrati	0.287	0.238
Solfati		0.873

1. Sapendo che le varianze specifiche di calcio e solfati sono rispettivamente pari a 0.489 e 0.115 si completi la tabella precedente;
2. si calcolino le varianze specifiche e la percentuale di varianza totale dovuta ai fattori specifici;
3. si interpreti sinteticamente il risultato ottenuto;
4. si calcoli la percentuale della variabilità comune complessiva dovuta al primo fattore latente.
5. Si descriva sinteticamente il metodo della regressione per la stima dei punteggi fattoriali.

Soluzione

1. Essendo per la i -esima variabile misurabile, $\sum_{t=1}^k \lambda_{it}^2 + \psi_i = 1$ si ha

per il calcio $\lambda_{12} = (1 - 0.453^2 - 0.489)^{1/2}$

per i solfati $\lambda_{61} = (1 - 0.873^2 - 0.115)^{1/2}$

2. per il magnesio $\psi_2 = 1 - 0.137^2 - 0.722^2 = 0.459$
per il sodio $\psi_3 = 0.043$
per i cloruri $\psi_4 = 0.267$
per i nitrati $\psi_5 = 0.861$

La somma delle varianze specifiche è pari a 2.235 quindi la quota di varianza totale (6) dovuta alla componente specifica è 37.3%

3. Il primo fattore è essenzialmente legato alle concentrazioni di Sodio e Cloruri.
Il secondo fattore è essenzialmente legato alle concentrazioni di Magnesio e Solfati.
I Nitrati sono poco spiegati dai fattori latenti (varianza specifica 86% della varianza della variabile).
Il modello spiegando il 73% circa della variabilità osservata tramite i fattori latenti risulta adeguato per i dati considerati.
4. Essendo la somma delle comunalità dovuta al primo fattore latente per le sei variabili pari a 2 si ha che la percentuale di comunalità complessiva ($6 - 2.235 = 3.765$) ad esso dovuta è circa il 53%. I due fattori sono rilevanti (circa) in egual misura nel rappresentare la componente spiegata dei dati.