

Correspondence Analysis

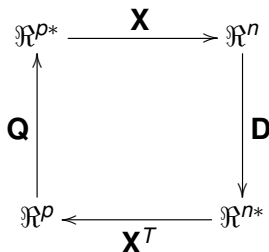
Analisi dei Dati¹

¹Corso di Laurea in Scienze Statistiche e Attuariali
Dipartimento di Diritto, Economia, Management e Metodi Quantitativi (DEMM)
Università degli Studi del Sannio

Prof. Pietro Amenta

Fonte: *Pietro Amenta. Appunti di Analisi dei Dati Multidimensionali*

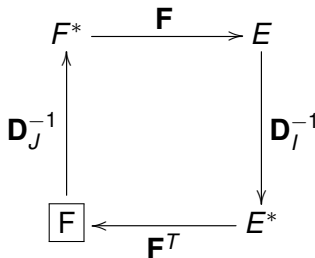
Abbiamo visto che se riportiamo le relazioni di una tripletta statistica $(\mathbf{X}, \mathbf{Q}, \mathbf{D})$ in uno schema grafico, otteniamo lo schema di base del **Diagramma di Dualità**. Tale schema evidenzia tutte le relazioni di una Analisi in Componenti Principali della matrice \mathbf{X} con metrica \mathbf{Q} e pesi \mathbf{D} : ACP($\mathbf{X}, \mathbf{Q}, \mathbf{D}$).



Possiamo ottenere le relazioni di base dell'AC utilizzando la tripletta statistica e il diagramma di dualità. Infatti l'AC risulta essere una particolare ACP.

AC come $\text{ACP}(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1})$

1. $\text{AC come } \text{ACP}(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1})$ Consideriamo lo schema di dualità della tripletta $(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1})$, con $\mathbf{u} \in F$



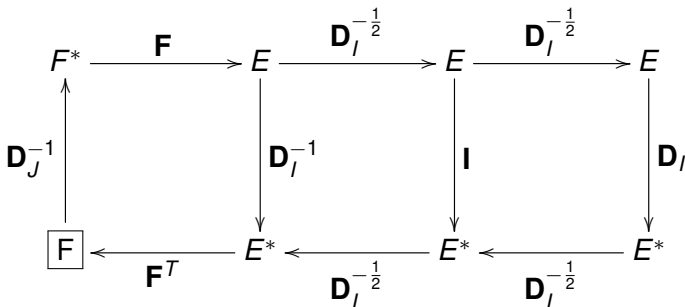
$$\text{ACP}(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1}) \rightarrow \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u} = \lambda \mathbf{u}$$

$$\mathbf{D}_J^{-1/2} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1/2} \mathbf{g} = \lambda \mathbf{g} \longleftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{-1/2} \mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{1/2} \mathbf{g}$.

AC come $ACP(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I)$

2. AC come $ACP(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I)$ Sviluppiamo il diagramma di dualità della tripletta $(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I)$, con $\mathbf{u} \in F$



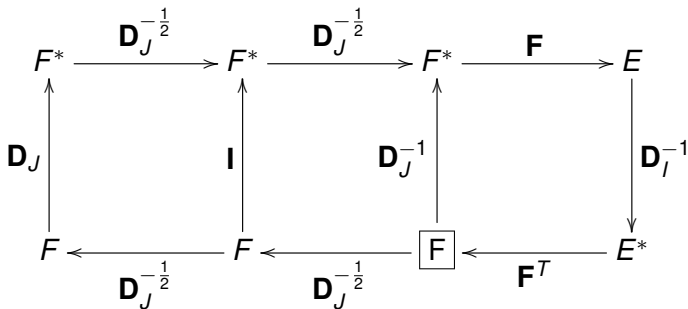
$$ACP(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I) \rightarrow \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{D}_I \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u} = \lambda \mathbf{u}$$

$$\mathbf{D}_J^{-1/2} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1/2} \mathbf{g} = \lambda \mathbf{g} \leftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{-1/2} \mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{1/2} \mathbf{g}$.

AC come $\text{ACP}(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})$

3. $\text{AC come } \text{ACP}(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})$ Consideriamo ora il diagramma di dualità della tripletta $(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})$



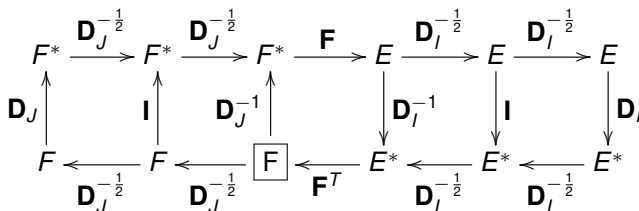
$$\boxed{\text{ACP}(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})} \rightarrow \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{u} = \lambda \mathbf{u}$$

$$\mathbf{D}_J^{-1/2} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1/2} \mathbf{g} = \lambda \mathbf{g} \leftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{1/2} \mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{-1/2} \mathbf{g}$.

AC come $ACP(\mathbf{D}_I^{-1}\mathbf{P}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I)$

4. AC come $ACP(\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I)$ Consideriamo infine il diagramma di dualità della tripletta $(\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I)$



$$ACP(\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I) \rightarrow \mathbf{D}_J^{-1}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{D}_I\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}\mathbf{D}_J\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1/2}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1/2}\mathbf{g} = \lambda\mathbf{g} \leftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{1/2}\mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{-1/2}\mathbf{g}$.

Consideriamo le formule di transizione dell'AC

$$\mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha$$

$$\mathbf{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha$$

Le coordinate delle modalità riga e colonna si possono scrivere

$$\begin{aligned} \psi_\alpha &= \mathbf{D}_I^{-1} \mathbf{F} \mathbf{u}_\alpha^* = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha \\ &= \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha = \sqrt{\lambda_\alpha} \mathbf{D}_I^{-1} \mathbf{v}_\alpha \\ \phi_\alpha &= \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{v}_\alpha^* = \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha = \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha \\ &= \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha = \sqrt{\lambda_\alpha} \mathbf{D}_J^{-1} \mathbf{u}_\alpha \end{aligned}$$

Consideriamo le coordinate riga $\psi_\alpha = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{u}_\alpha^*$ con $\mathbf{Q} = \mathbf{D}_J^{-1}$

$$\mathbf{D}_I^{-1} \mathbf{F} \mathbf{u}_\alpha^* = \sqrt{\lambda_\alpha} \mathbf{D}_I^{-1} \mathbf{v}_\alpha$$

$$\mathbf{D}_I^{-1} \mathbf{F} \mathbf{u}_\alpha^* \mathbf{u}_\alpha^{*T} \mathbf{Q}^{-1} = \sqrt{\lambda_\alpha} \mathbf{D}_I^{-1} \mathbf{v}_\alpha \mathbf{u}_\alpha^{*T} \mathbf{Q}^{-1}$$

$$\mathbf{D}_I^{-1} \mathbf{F} \sum_{\alpha} \mathbf{u}_\alpha^* \mathbf{u}_\alpha^{*T} \mathbf{Q}^{-1} = \sum_{\alpha} \sqrt{\lambda_\alpha} \mathbf{D}_I^{-1} \mathbf{v}_\alpha \mathbf{u}_\alpha^{*T} \mathbf{Q}^{-1}$$

$$\mathbf{D}_I^{-1} \mathbf{F} \sum_{\alpha} \mathbf{u}_\alpha^* \mathbf{u}_\alpha^{*T} \mathbf{D}_J = \sum_{\alpha} \sqrt{\lambda_\alpha} \mathbf{D}_I^{-1} \mathbf{v}_\alpha \mathbf{u}_\alpha^{*T} \mathbf{D}_J$$

$$\mathbf{D}_I^{-1} \mathbf{F} = \sum_{\alpha} \psi_\alpha \mathbf{u}_\alpha^{*T} \mathbf{D}_J$$

$$\mathbf{D}_I^{-1} \mathbf{F} = \sum_{\alpha} \frac{1}{\sqrt{\lambda_\alpha}} \psi_\alpha \phi_\alpha^T \mathbf{D}_J$$

dove $\sum_{\alpha} \mathbf{u}_\alpha^* \mathbf{u}_\alpha^{*T} \mathbf{D}_J = \mathbf{I}$ e $\phi_\alpha = \sqrt{\lambda_\alpha} \mathbf{u}_\alpha^*$

L'elemento generico della matrice $\mathbf{D}_I^{-1}\mathbf{F}$ si può allora scrivere

$$\frac{f_{ij}}{f_{i.}} = f_{.j} \sum_{\alpha} \frac{\psi_{i\alpha} \phi_{j\alpha}}{\sqrt{\lambda_{\alpha}}}$$

e quindi

$$f_{ij} = f_{i.} f_{.j} \sum_{\alpha} \frac{\psi_{i\alpha} \phi_{j\alpha}}{\sqrt{\lambda_{\alpha}}} = f_{i.} f_{.j} \left(\frac{\psi_{i1} \phi_{j1}}{\sqrt{\lambda_1}} + \sum_{\alpha > 1} \frac{\psi_{i\alpha} \phi_{j\alpha}}{\sqrt{\lambda_{\alpha}}} \right)$$

Si ricordi che $\lambda_1 = 1$ (autovalore banale) e che $\psi_{i1} = \phi_{j1} = 1 \forall(i, j)$ poichè \mathbf{u}_1 coincide con il baricentro (profilo medio $f_{.j}$)

$$\psi_1 = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_1 = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \begin{bmatrix} f_{.1} \\ \vdots \\ f_{.J} \end{bmatrix} = \mathbf{D}_I^{-1} \mathbf{F} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Possiamo quindi ricostruire l'elemento generico della matrice **F** attraverso le coordinate delle modalità riga e colonna

$$f_{ij} = f_{i.} f_{.j} \left(1 + \sum_{\alpha > 1} \frac{\psi_{i\alpha} \phi_{j\alpha}}{\sqrt{\lambda_{\alpha}}} \right)$$

Se invece consideriamo le le coordinate delle modalità riga e colonna normalizzate ad 1 $\hat{\psi}_{i\alpha} = \psi_{i\alpha} / \sqrt{\lambda_{\alpha}}$ e $\hat{\phi}_{j\alpha} = \phi_{j\alpha} / \sqrt{\lambda_{\alpha}}$, allora abbiamo che $\psi_{i\alpha} = \hat{\psi}_{i\alpha} \sqrt{\lambda_{\alpha}}$ e $\phi_{j\alpha} = \hat{\phi}_{j\alpha} \sqrt{\lambda_{\alpha}}$. Possiamo scrivere allora sostituendo

$$\begin{aligned} f_{ij} &= f_{i.} f_{.j} \left(1 + \sum_{\alpha > 1} \lambda_{\alpha} \frac{\hat{\psi}_{i\alpha} \hat{\phi}_{j\alpha}}{\sqrt{\lambda_{\alpha}}} \right) \\ &= f_{i.} f_{.j} \left(1 + \sum_{\alpha > 1} \sqrt{\lambda_{\alpha}} \hat{\psi}_{i\alpha} \hat{\phi}_{j\alpha} \right) \end{aligned}$$