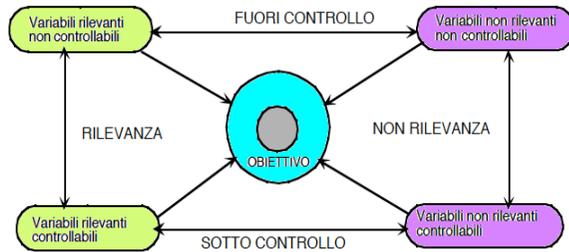


Legami tra variabili

In ogni problema ci sono fattori controllabili, incontrollabili ed altri sconosciuti, ma che non per questo evitano di influenzare sia la decisione che le conseguenze che da questa scaturiscono.

Alcuni fattori sono rilevanti ed altri no. Le variazioni nei primi comportano cambiamenti apprezzabili, negli altri non hanno effetti sensibili.

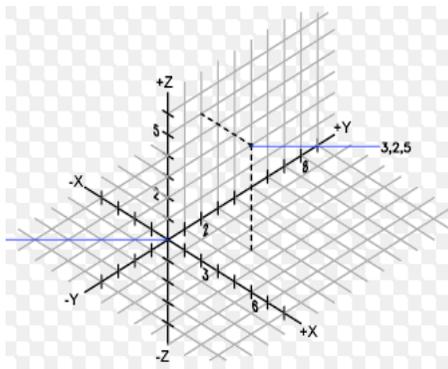


Il successo o l'insuccesso di una iniziativa dipende da come le variabili rilevanti interagiscono tra di esse, dalla relazione che le lega all'obiettivo e dalla concomitanza fortunata che i fattori latenti assecondino o non ostacolino o inibiscano il progresso delle linee di azione risolutive.

Variabili metriche

I legami diventano di vicinanza, contiguità, prossimità, dipendenza quando si riferiscono a variabili quantitative per il significato metrico e quindi di distanza che hanno qui le modalità.

Ogni unità può essere un punto in un sistema cartesiano di assi (formanti angoli retti) ed ogni idea di somiglianza tra di esse potrà essere espressa come vicinanza.

$$\begin{bmatrix}
 & X_1 & \cdots & X_m \\
 u_1 & x_{11} & \cdots & x_{1m} \\
 u_2 & x_{21} & \cdots & x_{2m} \\
 \vdots & & & \\
 u_i & x_{i1} & \ddots & x_{im} \\
 \vdots & & & \\
 u_n & x_{n1} & & x_{nm}
 \end{bmatrix}$$


La rappresentazione geometrica consente di ragionare in astratto per problemi più generali. I metodi ottenuti saranno applicati ai casi particolari.

Obiettivo: semplificare!

La disposizione dei punti dallo spazio è la struttura che deve essere semplificata e spiegata.

L'analisi multivariata può fare riferimento sia alle unità che alle variabili ed in base alla scelta si usano certe tecniche e non altre.

L'intento rimane comunque lo stesso: semplificare!

Riduzione del numero di unità: gli n distinti soggetti o entità sono ricondotti a pochi casi tipici le cui caratteristiche rappresentano gli aspetti salienti nel data set (ad esempio l'analisi dei gruppi)

Riduzione del numero di variabili: gli m distinti indicatori sono combinati in pochi altri fattori o variabili latenti che descrivono sinteticamente, al meglio possibile, il contenuto informativo del data set (ad esempio l'analisi delle componenti principali).

Rappresentazione geometrica dei dati

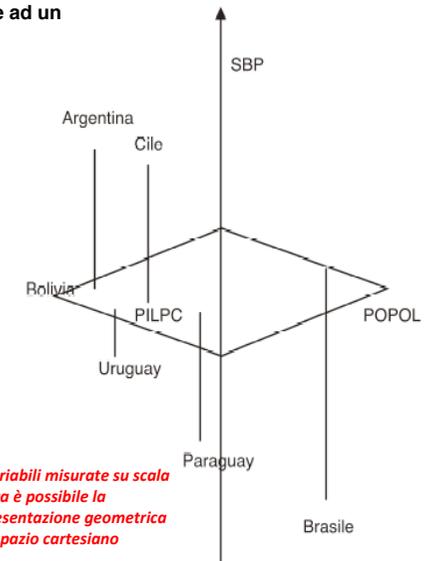
Ogni riga della matrice dei dati corrisponde ad un punto nello spazio R^m

$$(x_{i1}, x_{i2}, \dots, x_{im})$$

Sono incluse le rilevazioni sulle variabili relative alla i -esima unità. Di tali vettori se ne usano n (l'ampiezza della rilevazione). L'insieme delle rilevazioni costituisce la matrice dei dati.

	Pop.	Pil.pc	Sbp	
$X =$	34.6	8.121	2.165	Argentina
	155.9	4.335	-3.200	Brasile
	4.8	1.854	-1.675	Paraguay
	3.2	5.563	-0.761	Uruguay
	8.1	8.640	-0.222	Bolivia
	14.2	4.739	1.374	Cile

Per variabili misurate su scala metrica è possibile la rappresentazione geometrica nello spazio cartesiano



Rappresentazione geometrica dei dati/2

La rappresentazione in R^3 avviene a mezzo dei vettori coordinate formati da zero e da un solo "1" in una posizione specifica

L'Argentina è così rappresentata da

$$\begin{bmatrix} 34.6 \\ 8.121 \\ 2.165 \end{bmatrix} = 34.6 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 8.121 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 2.165 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}; \quad e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \text{l'uno è in posizione "i"}$$

Per varie ragioni, le informazioni qualitative sono espresse con numeri anche se sono effettuabili solo comparazioni del tipo uguali/diverse. In pratica, le modalità sono codificate per semplificarne le elaborazioni.

Nel caso in esempio però le informazioni sono tutte su scala a rapporti o intervallare cioè si possono assimilare a dei punti nello spazio.

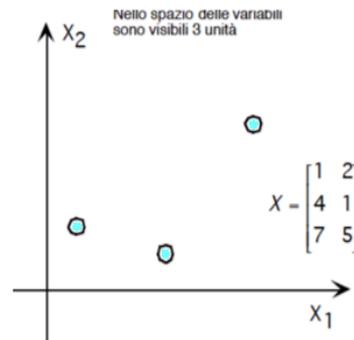
Spazio delle variabili

In alternativa si potrebbe ragionare sulle relazioni tra le unità ed in questo caso la matrice dei dati è considerata un aggregato di "n" righe.

$$X = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad u_i = [u_{i1}, u_{i2}, \dots, u_{im}]$$

Ogni u_i è il vettore riga formato dalle rilevazioni effettuate su tutte le m variabili rispetto alla i-ma unità.

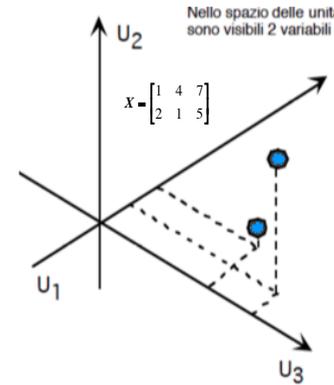
Si dice che in queste analisi ci si muove nello spazio delle variabili, sottinsieme di R^m .



Spazio delle unità

Le analisi statistiche della matrice dei dati sono talvolta dirette alle relazioni tra le variabili ed in questo caso la si considera un aggregato di "m" colonne

$$X = [x_1, x_2, \dots, x_m] \quad x_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix}$$



Ogni x_i è un vettore colonna formato dalle osservazioni per la i-esima variabile su tutte le n unità coinvolte nella rilevazione.

Si dice che in questa analisi ci si muove nello spazio delle unità che è un sottinsieme di R^n .

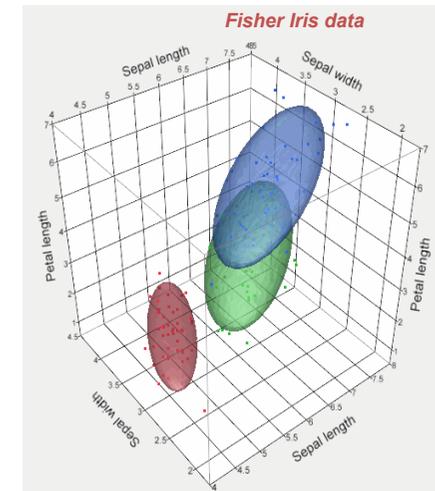
Qui, ogni riga corrisponde ad un punto.

Nube dei punti

La disposizione dei punti dallo spazio (dalle variabili o delle unità) è la struttura che deve essere spiegata e semplificata.

L'analisi multivariata può fare riferimento sia al numero di unità n che al numero di variabili m.

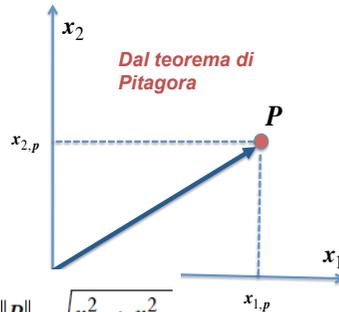
Un punto in comune è che, sia nel caso delle unità che nel caso delle variabili, le dimensioni effettivamente in gioco non sono n o m, ma dei valori inferiori e, sperabilmente, molto inferiori.



Norma euclidea di un vettore

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^m x_i^2}$$

Come mostra la figura, un vettore può essere rappresentato come una freccia che si diparte dall'origine per raggiungere un punto le cui coordinate sono gli elementi del vettore



\overline{OP} è il vettore $P = \begin{bmatrix} x_{1P} \\ x_{2P} \end{bmatrix}$ con norma $\|P\| = \sqrt{x_{1P}^2 + x_{2P}^2}$

$$P = x_{1P} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_{2P} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

I vettori si dicono normalizzati se la loro norma euclidea è pari ad uno

$$\mathbf{x}^* = \frac{x_i}{\|\mathbf{x}\|} \quad i=1,2,\dots,n \Rightarrow \|\mathbf{x}^*\| = \sqrt{\sum_{i=1}^m \left(\frac{x_i}{\|\mathbf{x}\|}\right)^2} = \left(\frac{1}{\|\mathbf{x}\|}\right) \sqrt{\sum_{i=1}^m x_i^2} = \left(\frac{1}{\|\mathbf{x}\|}\right) \|\mathbf{x}\| = 1$$

Legge dei coseni (teorema di Carnot)/2

Ne consegue che

$$\begin{aligned} c^2 &= d^2 + (b-x)^2 = d^2 + b^2 - 2bx + x^2 \\ &= a^2 \sin^2 \theta + b^2 - 2ab \cos \theta + a^2 \cos^2 \theta \\ &= a^2 [\sin^2 \theta + \cos^2 \theta] + b^2 - 2ab \cos \theta \end{aligned}$$

Poiché $\sin^2 \theta + \cos^2 \theta = 1$ si ha $c^2 = a^2 + b^2 - 2ab \cos \theta$

Legge di Carnot

Ricorrendo ai vettori si ottiene:

$$\begin{aligned} c &= \|P_1 - P_2\| \\ a &= \|P_1\|; \quad b = \|P_2\| \end{aligned}$$

E quindi

$$\|P_1 - P_2\|^2 = \|P_1\|^2 + \|P_2\|^2 - 2\|P_1\| \|P_2\| \cos \theta$$

Legge dei coseni (teorema di Carnot)

I vettori posso differire per lunghezza, per l'angolo formato con gli assi o per entrambi.

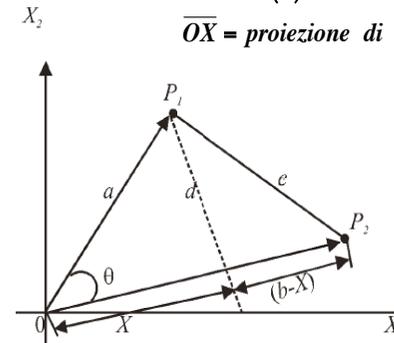
Rispetto all'angolo θ possiamo stabilire un utile teorema.

$$d = a \sin(\theta)$$

$$x = a \cos(\theta)$$

$\overline{OX} =$ proiezione di P_1 su P_2

d può essere visto come la distanza perpendicolare tra P_1 e P_2



Se si conosce l'angolo θ formato dai due vettori, la proiezione di un vettore sulla retta dell'altro si ottiene moltiplicando il suo modulo per il coseno di θ .

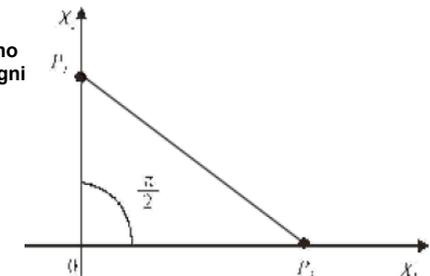
Esempio

Se i due vettori sono ortogonali e cioè se $\theta = \frac{\pi}{2} \Rightarrow 90^\circ$

allora $\cos(\theta)=0$ e la distanza tra i due vettori è

$$\|P_1 - P_2\|^2 = \|P_1\|^2 + \|P_2\|^2$$

Questo succede quando i punti si trovano sugli assi di un sistema cartesiano ed ogni punto su di un diverso asse



Prodotto scalare (o interno)

Consideriamo due colonne della matrice dei dati

$$X_1' = (x_{1,1}, x_{1,2}, \dots, x_{1,n}); \quad X_2' = (x_{2,1}, x_{2,2}, \dots, x_{2,n})$$

Il loro prodotto scalare (in \mathbb{R}^n) è il numero reale ottenuto dalla somma dei prodotti delle componenti omologhe, ossia

$$X_1' X_2 = \sum_{h=1}^n x_{h,1} x_{h,2}$$

Esempio $\begin{bmatrix} X_1 & X_2 \\ 0 & 3 \\ 1 & 1 \\ -1 & 2 \end{bmatrix}; X_1' X_2 = 0 \cdot 3 + 1 \cdot 1 + (-1) \cdot 2 = -1$

Il prodotto scalare (o interno) di due vettori dipende dalla lunghezza dei vettori (cioè dalla loro norma) e dall'angolo che essi formano

Interpretazione del prodotto interno/2

Si ha

$$-2(x_{11} - x_{21} + x_{12} - x_{22}) = -2\overline{OP_1} \cdot \overline{OP_2} \cos \theta$$

$$(x_{11} - x_{21} + x_{12} - x_{22}) = \left[\sqrt{(x_{11}^2 + x_{21}^2)} \cdot \sqrt{(x_{12}^2 + x_{22}^2)} \right] \cdot \cos \theta$$

E quindi

$$\cos \theta = \frac{x_{11}x_{21} + x_{12}x_{22}}{\sqrt{(x_{11}^2 + x_{21}^2)} \cdot \sqrt{(x_{12}^2 + x_{22}^2)}} = \frac{P_1' P_2}{\|P_1\| \cdot \|P_2\|} = \frac{P_1'}{\|P_1\|} \cdot \frac{P_2}{\|P_2\|}$$

Il prodotto scalare è pari al prodotto delle lunghezze dei due vettori per il coseno del loro angolo.

$$P_1' P_2 = \|P_1\| \cdot \|P_2\| \cos \theta$$

Se i due vettori sono normalizzati, il loro prodotto scalare è pari al coseno del loro angolo

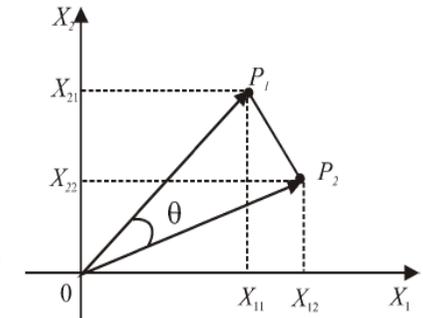
$$P_1^+ P_2^+ = \cos \theta \Rightarrow -1 \leq P_1^+ P_2^+ \leq 1$$

Interpretazione del prodotto interno

$$P_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}; \quad P_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$$

Dal teorema di Carnot sappiamo che

$$\overline{P_1 P_2}^2 = \overline{OP_1}^2 + \overline{OP_2}^2 - 2\overline{OP_1} \cdot \overline{OP_2} \cos \theta$$



I segmenti possono essere espressi come distanze euclidee (al quadrato)

$$\overline{P_1 P_2}^2 = (x_{11}^2 - x_{21}^2)^2 + (x_{12}^2 - x_{22}^2)^2 = (x_{11}^2 + x_{21}^2) + (x_{12}^2 + x_{22}^2) - 2(x_{11}x_{21} + x_{12}x_{22})$$

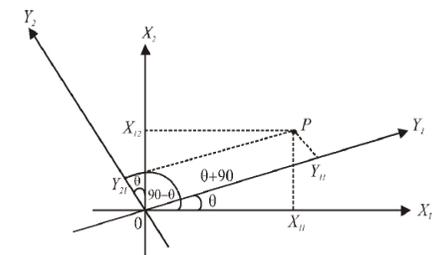
Poiché

$$\overline{OP_1}^2 = (x_{11}^2 + x_{12}^2); \quad \overline{OP_2}^2 = (x_{21}^2 + x_{22}^2)$$

Rotazioni degli assi

È spesso necessaria o opportuna la trasformazione del piano cartesiano: rotazioni, contrazioni, dilatazioni, deformazioni, etc. allo scopo di semplificare il problema trattato.

Ogni trasformazione può essere descritta con delle equazioni che collegano il piano cartesiano originale con il nuovo sistema di coordinate.



Il vettore P può essere scritto in entrambi i sistemi come:

sistema (x_1, x_2)	sistema (y_1, y_2)
$P = x_{11}e_1 + x_{12}e_2$	$P = y_{11}e_1^* + y_{12}e_2^*$

dove e_1, e_2, e_1^*, e_2^* sono dei vettori normalizzati ed ortogonali (ovvero vettori coordinate) tali che

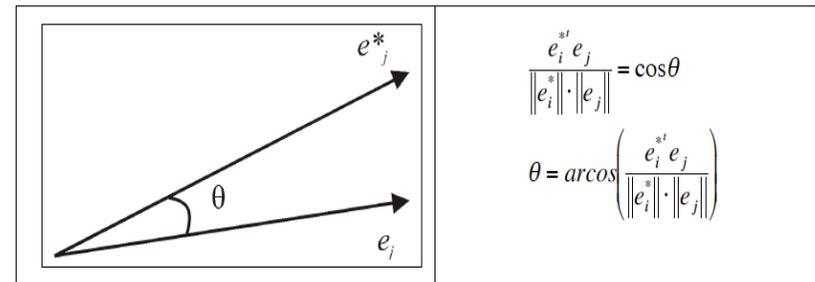
$$e_i^T e_j = d_{ij} \quad e_i^* T e_j^* = d_{ij} \quad d_{ij} = \text{delta di Kronecker} \Rightarrow d_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Rotazione degli assi/3

Cerchiamo di scoprire la natura geometrica del particolare prodotto scalare $e_i^{*t} e_j$. Innanzitutto:

$$e_i^{*t} e_j = \frac{e_i^{*t} e_j}{\|e_i^*\| \cdot \|e_j\|} \quad \text{dato che } \|e_i^*\| = \|e_j\| = 1$$

Il rapporto tra il prodotto scalare ed il prodotto delle norme dei vettori, come si è visto, è pari al coseno dell'angolo tra i due vettori



Nel sistema (x_1, x_2) effettuiamo il prodotto scalare: **Rotazione degli assi/2**

$$e_1^{*t} (x_{11} \cdot e_1 + x_{12} \cdot e_2) = e_1^{*t} P = x_{11} e_1^{*t} e_2 + x_{12} e_1^{*t} e_2$$

Lo ripetiamo nel sistema (y_1, y_2) per ottenere:

$$e_1^{*t} (y_{11} \cdot e_1^* + y_{12} \cdot e_2^*) = e_1^{*t} P = y_{11} e_1^{*t} e_1^* + y_{12} e_1^{*t} e_2^* = y_{11} \cdot 1 + y_{12} \cdot 0 = y_{11}$$

e quindi, eguagliando i due risultati, si avrà

$$x_{11} e_1^{*t} e_1 + x_{12} e_1^{*t} e_2 = y_{11}$$

Allo stesso modo, effettuando il prodotto scalare per il secondo vettore per P, si ottiene

$$e_2^{*t} P = x_{11} e_2^{*t} e_1 + x_{12} e_2^{*t} e_2 ;$$

Si arriva pertanto al sistema:

$$y_{11} = x_{11} e_1^{*t} e_1 + x_{12} e_1^{*t} e_2$$

$$y_{12} = x_{11} e_2^{*t} e_1 + x_{12} e_2^{*t} e_2$$

Rotazione degli assi/4

Ad ogni prodotto scalare del sistema di trasformazione può essere applicata la relazione con il coseno (tenuto conto del giusto angolo)

$$e_1^{*t} e_1 = \cos \theta ; \quad e_2^{*t} e_1 = \cos(\theta + 90) = -\sin(\theta)$$

$$e_1^{*t} e_2 = \cos(90 - \theta) = \sin \theta ; \quad e_2^{*t} e_2 = \cos \theta$$

Il sistema di trasformazione può quindi essere riespresso come

$$y_{11} = x_{11} \cos \theta - x_{12} \sin \theta$$

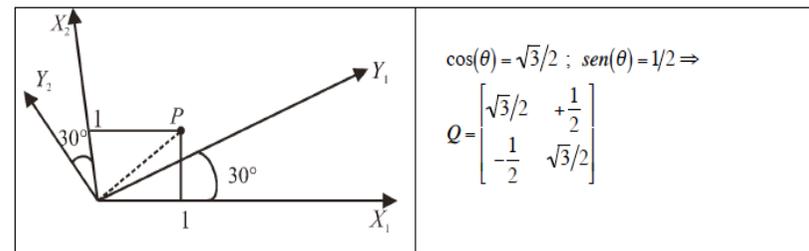
$$y_{12} = x_{11} \sin \theta + x_{12} \cos \theta$$

$$\begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} \cos \theta & +\sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} = x_{11} \begin{bmatrix} \cos \theta \\ -\sin \theta \end{bmatrix} + x_{12} \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix}$$

In notazione matriciale compatta si ha: $\mathbf{Y} = \mathbf{XQ}$ dove \mathbf{Q} è una matrice ortogonale detta di rotazione. Tale matrice è anche ortonormale cioè ortogonale e con norma unitaria di righe e di colonne.

Esempio

Supponiamo che $\theta = 30^\circ$ e determiniamo la corrispondente matrice di rotazione



$$\cos(\theta) = \sqrt{3}/2 ; \quad \sin(\theta) = 1/2 \Rightarrow$$

$$\mathbf{Q} = \begin{bmatrix} \sqrt{3}/2 & +1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}$$

Il punto P rimane fermo. È il sistema degli assi che ruota. $P=(1,1)$ in (X_1, X_2) in (Y_1, Y_2) diventa

$$\begin{bmatrix} \sqrt{3}/2 & +1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (\sqrt{3}+1)/2 \\ (\sqrt{3}-1)/2 \end{bmatrix}$$

Se nel vecchio sistema, le coordinate del punto avevano lo stesso valore, nel nuovo sistema i rapporti cambiano e una delle coordinate diventa maggiore dell'altra.

Questo, come vedremo, non è privo di significato.

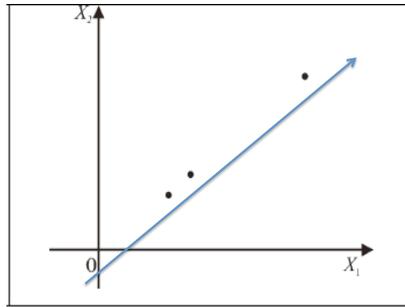
Riduzione delle dimensioni

Consideriamo i punti:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}; \begin{bmatrix} 2 \\ 2 \end{bmatrix}; \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

I punti sono perfettamente allineati lungo la retta $x_2=x_1$. Nota una coordinata l'altra si determina in modo automatico.

Si vuole ora spostare l'asse x_1 in modo da farlo coincidere con la retta di equazione $x_2=x_1$



La matrice di rotazione necessaria per tale operazione è.

$$\theta = \frac{\pi}{4} \Rightarrow \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & +\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

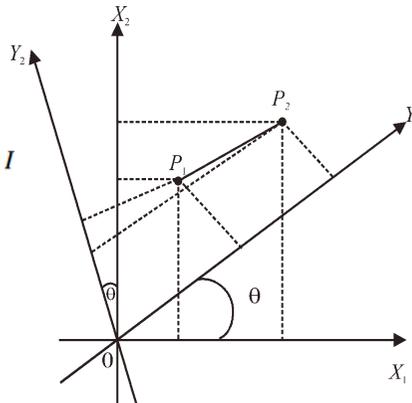
Rotazione degli assi e distanze

Le rotazioni ortogonali mantengono inalterate le distanze tra i punti.

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \Rightarrow R^t R = R R^t = I$$

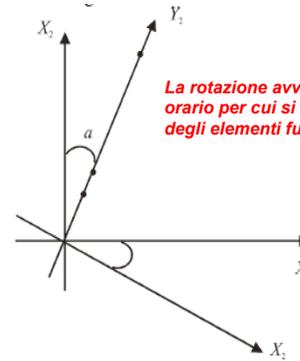
Indichiamo i nuovi punti nel sistema ruotato con

$$Q_1 = R P_1 \quad \text{e} \quad Q_2 = R P_2$$



$$\begin{aligned} \|Q_1 - Q_2\|^2 &= (Q_1 - Q_2)^t (Q_1 - Q_2) = Q_1^t Q_1 - Q_1^t Q_2 - Q_2^t Q_1 + Q_2^t Q_2 \\ &= P_1^t (R^t R) P_1 - P_1^t (R^t R) P_2 - P_2^t (R^t R) P_1 + P_2^t (R^t R) P_2 \\ &= P_1^t P_1 - P_1^t P_2 - P_2^t P_1 + P_2^t P_2 = (P_1 - P_2)^t (P_1 - P_2) = \|P_1 - P_2\|^2 \end{aligned}$$

Riduzione delle dimensioni/2



La rotazione avviene in senso orario per cui si inverte il segno degli elementi fuori diagonale.

I punti trasformati sono

$$\frac{1}{2} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$

$$\frac{1}{2} \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2\sqrt{2} \end{bmatrix}$$

$$\frac{1}{2} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 6\sqrt{2} \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 2\sqrt{2} \end{bmatrix}$$

Originali

Nelle nuove coordinate

$$\begin{bmatrix} 1 & 2 & 6 \\ 1 & 2 & 6 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ \sqrt{2} & 2\sqrt{2} & 6\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 6 \end{bmatrix}$$

Di fatto è stata eliminata una dimensione che era solo un duplicato dell'altra: basta un solo asse per rappresentare i tre punti dati.

Nella realtà in punti non sono mai perfettamente allineati. Comunque una opportuna rotazione degli assi consente di trascurare dimensioni poco informative.

Rotazione della matrice dei dati

Notiamo subito che la distanza tra i punti rimane invariata se le coordinate originali sono moltiplicate per una matrice di rotazione ortogonale per cui

$$\begin{aligned} Y &= X Q^t \Rightarrow Y Y^t = (X Q^t) (X Q^t)^t \\ &= X Q^t Q X^t \\ &= X X^t \end{aligned}$$

La matrice dei dati in Y è equivalente ad X, per ogni matrice di rotazione Q

L'angolo di rotazione è scelto qui arbitrariamente, ma esistono tecniche per una scelta ottimale in un qualche senso definito.

E' anomala quella unità che ha grande distanza dalla quasi totalità delle altre.

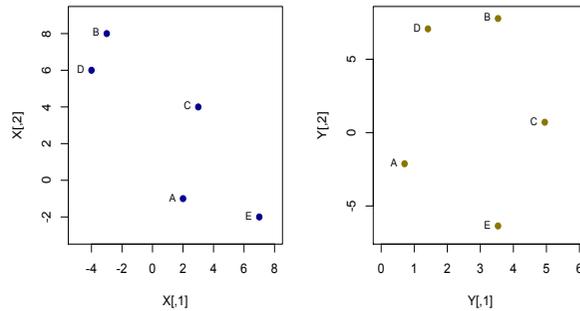
Attenzione! Se esistono dei valori anomali nella matrice dei dati, le anomalie permangono anche dopo la rotazione poiché le distanze non si modificano

Esempio

$$X = \begin{bmatrix} 2 & -1 \\ -3 & 8 \\ 3 & 4 \\ -4 & 6 \\ 7 & -2 \end{bmatrix}; \theta = 45^\circ \Rightarrow Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

$$XQ^t = \begin{bmatrix} 2 & -1 \\ -3 & 8 \\ 3 & 4 \\ -4 & 6 \\ 7 & -2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 0.707 & -2.121 \\ 3.536 & 7.778 \\ 4.950 & 0.707 \\ 1.414 & 7.071 \\ 3.536 & -6.364 \end{bmatrix}$$

Due erano le variabili originali e due sono le trasformate. Possiamo fare di meglio.

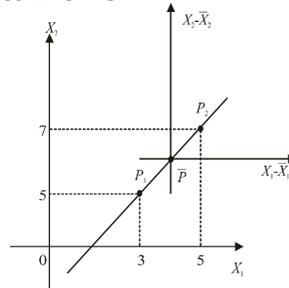


Scelta del perno di rotazione

La rotazione non basta ad alterare il sistema dei pesi. Infatti, per i punti P1=(3,5) e P2=(5,7) si ha:

$$\begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 2 & -2 \\ \sqrt{2} & \sqrt{2} \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 4\sqrt{2} & 6\sqrt{2} \end{bmatrix}$$

I punti sono allineanti, ma la semplice rotazione intorno all'origine non lo evidenzia.



Per superare questo problema, gli assi debbono essere spostati sul punto centrale tra quelli dati: P=(4,6) ottenuto come media aritmetica dei valori

$$\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 0 & 0 \end{bmatrix}$$

Così si chiarisce il numero vero di dimensioni necessarie per rappresentare i punti dati nell'esempio (una e non due come sembrava all'inizio).

Il centramento dei punti (misurazione come scarto dalle medie) consente di individuare le relazioni lineari perfette azzerando una o più dimensioni.

Variazione delle coordinate

L'uguaglianza delle norme – sotto rotazione ortogonale – non implica che gli elementi dei vettori abbiano lo stesso peso all'interno dei vettori ruotati.

Consideriamo il punto P=(2,√3) ed effettuiamo una rotazione degli assi di 30° (π/6)

$$\begin{bmatrix} \sqrt{3} & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ \sqrt{3} \end{bmatrix} = \begin{bmatrix} \sqrt{3} \\ 2 \\ 5 \\ 2 \end{bmatrix}$$

Sistema (x ₁ , x ₂)	Sistema (y ₁ , y ₂)
Nei due sistemi, la norma al quadrato nonché i pesi degli elementi (nell'ambito del vettore normalizzato) sono ora	
$\left(\frac{\sqrt{3}}{2}\right)^2 + \left(\frac{5}{2}\right)^2 = 1$	$(2)^2 + (\sqrt{3})^2 = 1$
$\frac{3}{28} + \frac{25}{28} = 1$	$\frac{16}{28} + \frac{12}{28} = 1$

Nel secondo sistema il peso delle dimensioni si inverte e l'asse più rilevante o preferenziale è y₁ e non più x₂.

È chiaro che a seconda della posizione del punto e dell'angolo di rotazione del sistema si può fare in modo da preconstituire l'ordinamento dei pesi desiderato.

Centramento e rotazione della matrice dei dati

Abbiamo introdotto due operazioni che possono agevolare la interpretazione della matrice dei dati: centramento e rotazione

$$X = \begin{bmatrix} 8 & 1 & 9 \\ 2 & 0 & 4 \\ 7 & -1 & 10 \\ 3 & 0 & 5 \end{bmatrix}$$

Supponiamo di ruotare X₁ ed X₂ di 30° rispetto ad X₃; X₁ ed X₃ di 60° rispetto ad X₂ e X₂ ed X₃ di 120° rispetto a X₁

$$Q_{x_3}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix} \Rightarrow Q_{x_3}(\pi/6) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & 0.5 \\ 0 & -0.5 & \sqrt{3}/2 \end{bmatrix} \text{ (asse Z)}$$

$$Q_{x_2}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \Rightarrow Q_{x_2}(\pi/3) = \begin{bmatrix} 0.5 & 0 & -\sqrt{3}/2 \\ 0 & 1 & 0 \\ \sqrt{3}/2 & 0 & 0.5 \end{bmatrix} \text{ (asse Y)}$$

La matrice di rotazione finale risulta dal prodotto delle singole matrici di rotazione

$$Q_{x_1}(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow Q_{x_1}(2\pi/3) = \begin{bmatrix} -0.5 & \sqrt{3}/2 & 0 \\ -\sqrt{3}/2 & -0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ (asse X)}$$

$$R = Q_{x_3} Q_{x_2} Q_{x_1} = \begin{bmatrix} -0.2500000 & -0.4330127 & 0.8660254 \\ 0.5334936 & -0.8080127 & -0.2500000 \\ 0.8080127 & 0.3995191 & 0.4330127 \end{bmatrix}$$

```
> C%*%A%*%t(R)
      [,1] [,2] [,3]
[1,] 0.5490381 0.2924682 3.689583
[2,] -1.8480762 -0.8504809 -3.723076
[3,] 2.5310889 1.1250000 2.515544
[4,] -1.2320508 -0.5669873 -2.482051
```

L'esito del centramento e della rotazione non è eccellente, ma può essere perfezionato

Notazione generale

In termini generali il sistema di trasformazioni può essere scritto come

$$Y = XQ^t$$

dove Q è una matrice di rotazione. Tale matrice è anche orto-normale cioè ortogonale e con norma unitaria di righe e di colonne.

Quando ci capita di osservare un oggetto sconosciuto è naturale ruotarlo in tanti modi per guardarlo da più punti di vista.

Spesso troviamo un'angolazione più comoda di altre secondo la quale l'oggetto ci appare più familiare o, almeno, possiamo descriverlo in termini più semplici.

$$Q^t Q = Q Q^t = \begin{bmatrix} \cos\theta & -\sin\theta \\ +\sin\theta & \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};$$

Diverse tecniche statistiche multivariate si basano su questo principio.

Combinazioni lineari

L'idea guida della analisi delle componenti principali (ACP) è cercare il numero minimo di trasformate lineari delle variabili originarie che conservino la massima informazione sulle correlazioni originarie.

$$F_i = \sum_{j=1}^m q_{ij} X_j, \quad i = 1, 2, \dots, p$$

Più che un problema statistico è un problema di algebra lineare da risolvere con il calcolo numerico.

La ACP trasforma le variabili X_1, X_2, \dots, X_m in un nuovo insieme di variabili ortogonali F_1, F_2, \dots, F_p dette componenti, tali che p sia molto più piccolo di m ed ogni componente F sia una combinazione lineare delle X .

$$\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1m} \\ q_{21} & q_{22} & \dots & q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{p1} & q_{p2} & \dots & q_{pm} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$$

Analisi delle componenti principali

In una indagine ci si potrebbe trovare di fronte alla seguente matrice di correlazione

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	.8	.9	.8	.2	.1	.1	.1	.0
x_2		1	.9	.7	.1	.2	.1	.1	.2
x_3			1	.9	-.1	.1	.3	-.0	.0
x_4				1	.2	.1	-.1	.1	.2
x_5					1	.9	.1	.2	.1
x_6						1	.2	.3	.0
x_7							1	.9	-.8
x_8								1	-.9
x_9									1

Quante sono in realtà le variabili?

Alcune sono molto correlate tra di loro e poco correlate con altre. Ad esempio:

$\{X_1, X_2, X_3, X_4\}$, $\{X_5, X_6\}$, $\{X_7, X_8, X_9\}$.

Se fossimo costretti a scegliere un numero estremamente ridotto di variabili, su quali cadrebbe la nostra scelta?

Più che variabili però ci interessano le relazioni tra di esse. Quante relazioni effettive tra variabili esistono nella matrice? La risposta è facile: tre ovvero una per ogni gruppo di variabili.

Qualche informazione va persa perché non si può pervenire a relazioni esatte, ma il risparmio sui dati sarebbe notevole: 3 variabili rispetto a 9.

Riflessione sui pesi della combinazione

Gli elementi q_{ij} della combinazione sono incogniti. Per agevolare il loro calcolo è meglio centrare i dati e ragionare sulle rotazioni della matrice dei dati.

$$F_i = \sum_{j=1}^m q_{ij} \hat{X}_j, \quad i = 1, 2, \dots, p \Rightarrow F = \hat{X}Q^t$$

\hat{X}_j = variabile originaria centrata

F_i = variabile artificiale ottenuta con le rotazioni

I pesi della combinazione debbono essere numeri reali tali da rientrare in una matrice di rotazione.

Questo dovrebbe agevolare la comprensione delle strutture presenti nella matrice dei dati.

Ogni componente o nuova variabile deriva dal prodotto esterno delle vecchie variabili centrate con una riga della matrice di rotazione.

Componenti e prodotti esterni

$$\widehat{X}Q^t = [\widehat{x}_1 \quad \dots \quad \widehat{x}_m] \begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_p \end{bmatrix} \Rightarrow \mathbf{F}_i = \begin{bmatrix} \mathbf{q}_1 \widehat{x}_1 \\ \vdots \\ \mathbf{q}_p \widehat{x}_m \end{bmatrix}$$

In pratica si deve scegliere una nuova base per rappresentare i vettori, possibilmente in uno spazio con meno coordinate.

Il problema è proprio la scelta della matrice di rotazione Q.

Tale matrice deve rispondere all'obiettivo di azzerare (o quasi) quante più dimensioni possibili in modo da produrre la massima riduzione della dimensionalità senza disperdere informazioni rilevanti.

Da sottolineare l'uso dell'aggettivo lineare che induce una notevole semplificazione della ricerca di Q e quindi dei nuovi assi o componenti.

Scopo della ACP

L'idea della ACP parte da Galton che nel 1869 osservò che sfruttando le relazioni lineari o quasi-lineari tra indicatori si potesse ridurre le dimensioni dei dati arrivando ad una rappresentazione più parsimoniosa e più comprensibile

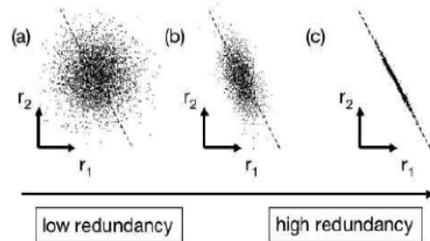
Le componenti "riassumono al meglio il contenuto informativo" delle variabili originarie.

Cosa vuol dire "riassumere al meglio"?

Le componenti riproducono al massimo la variabilità degli indicatori originari.

Perché?

Nella totale uniformità non ci sono valutazioni da fare o decisioni da prendere.



Nel grafico (c) è evidente la relazione lineare esatta tra le due variabili. Nel grafico (a) è difficile estrarre qualcosa; il caso più frequente è il (b).

Prodotto esterno di matrici

Sarà più comodo seguire le formule se si pensa al prodotto tra matrici realizzato come somma di prodotti colonne righe

$$= \begin{bmatrix} 7 & 2 & 5 & 9 \\ -1 & 4 & 1 & 8 \\ 3 & 6 & 0 & -3 \end{bmatrix}; B = \begin{bmatrix} -4 & -9 \\ 6 & 5 \\ 5 & 8 \\ 2 & -1 \end{bmatrix}; AB = \begin{bmatrix} 27 & -22 \\ 49 & 29 \\ 18 & 6 \end{bmatrix}$$

$$AB = \begin{bmatrix} 7 \\ -1 \\ 3 \end{bmatrix} \begin{bmatrix} -4 & -9 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \begin{bmatrix} 6 & 5 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 5 & 8 \end{bmatrix} + \begin{bmatrix} 9 \\ 8 \\ -3 \end{bmatrix} \begin{bmatrix} 2 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} -28 & -63 \\ 4 & 9 \\ -12 & -27 \end{bmatrix} + \begin{bmatrix} 12 & 10 \\ 24 & 20 \\ 36 & 30 \end{bmatrix} + \begin{bmatrix} 25 & 40 \\ 5 & 8 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 18 & -9 \\ 16 & -8 \\ -6 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 27 & -22 \\ 49 & 29 \\ 18 & 6 \end{bmatrix}$$

In questo modo il prodotto di matrici AB è dato dalla somma dei prodotti delle colonne di A per le righe di B.

Rapporti di variabilità

I dati reali sono contaminati da difetti di misurazione, imprecisioni, incertezze che con termine comune potremmo denominare come errori.

L'errore oscura la percezione e la determinazione di ciò che di valido e utile c'è nei dati.

Non esiste una scala assoluta per misurare l'errore, anche perché non si conoscono le ragioni e l'ammontare dell'errore, altrimenti sarebbe facile eliminarlo. Per ragionare ci si basa sul rapporto:

$$\frac{\text{Variabilità spiegata}}{\text{Variabilità totale}}$$

Maggiore è tale rapporto, più grande sarà la quantità valida di variabilità che si è potuto ottenere dai dati.

Se il rapporto è molto basso, la qualità dei dati che si è acquisita dimostra che c'era poco da acquisire oppure che la tecnica di acquisizione è inadatta.

Variabilità nella matrice dei dati

Il contenuto di variabilità di una matrice dei dati confluisce nella matrice di varianze e covarianze.

$$\left(\frac{1}{\sqrt{n}}\hat{\mathbf{X}}\right)\left(\frac{1}{\sqrt{n}}\hat{\mathbf{X}}\right)' = \left(\frac{1}{n^2}\right)\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X} = \left(\frac{1}{n^2}\right)\mathbf{X}'\mathbf{C}\mathbf{X} = \mathbf{W}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix}; \mathbf{C} = \begin{bmatrix} 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & -1/4 & 3/4 \end{bmatrix}$$

$$\hat{\mathbf{X}} = \begin{bmatrix} 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & -1/4 & 3/4 \end{bmatrix} \begin{bmatrix} 1 & 1 & -3 \\ 4 & 0 & 10 \\ 2 & 2 & 5 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -6 \\ 2 & -1 & 7 \\ 0 & 1 & 2 \\ -1 & 0 & -3 \end{bmatrix}$$

$$\hat{\mathbf{X}}'\hat{\mathbf{X}} = \begin{bmatrix} 6 & -2 & 23 \\ -2 & 2 & -5 \\ 23 & -5 & 98 \end{bmatrix}$$

$$= \begin{bmatrix} 6/4 & -2/4 & 23/4 \\ -2/4 & 2/4 & -5/4 \\ 23/4 & -5/4 & 98/4 \end{bmatrix} = \mathbf{W}$$

Ricerca della matrice di rotazione

L'analisi delle componenti principali consiste nel ruotare la matrice dei dati scegliendo gli angoli di rotazione uno ad uno e sempre orientati nella direzione della massima variabilità dei dati.

$$\mathbf{F}_1 = \hat{\mathbf{X}}\mathbf{q}_1 \quad \text{dove } \mathbf{q}_1 = (q_{11}, q_{12}, \dots, q_{1m})'$$

La variabilità intorno al nuovo asse farà parte della variabilità spiegata.

La scelta del vettore dei pesi implica la scelta degli angoli tra i vecchi assi ed il nuovo asse \mathbf{F}_1

Ogni scelta di angoli produce una diversa allocazione dei punti, ma noi siamo soprattutto interessati alla loro dispersione per cui ruoteremo il piano nella direzione in cui è massima la variabilità rispetto ad \mathbf{F}_1 .

Qual è la dispersione dei punti rispetto ad \mathbf{F}_1 ? Cioè, qual è la varianza di \mathbf{F}_1 ?

Variabilità nella matrice dei dati/2

La matrice di varianze-covarianze può essere calcolata anche come somma di prodotti esterni

$$\mathbf{W} = \frac{1}{n} \sum_{h=1}^n \hat{\mathbf{X}}_h \hat{\mathbf{X}}_h'$$

$$4\mathbf{W} = \begin{bmatrix} -1 \\ 0 \\ -6 \end{bmatrix} \begin{bmatrix} -1 & 0 & -6 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ 7 \end{bmatrix} \begin{bmatrix} 2 & -1 & 7 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ -3 \end{bmatrix} \begin{bmatrix} -1 & 0 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 6 \\ 0 & 0 & 0 \\ 6 & 0 & 36 \end{bmatrix} + \begin{bmatrix} 4 & -2 & 14 \\ -2 & 1 & -7 \\ 14 & -7 & 49 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 3 & 0 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & -2 & 23 \\ -2 & 2 & -5 \\ 23 & -5 & 98 \end{bmatrix} \Rightarrow \mathbf{W} = \begin{bmatrix} 6/4 & -2/4 & 23/4 \\ -2/4 & 2/4 & -5/4 \\ 23/4 & -5/4 & 98/4 \end{bmatrix}$$

Da notare che tutte le matrici addendi sono simmetriche e di rango uno.

Se si esclude uno dei prodotti esterni cioè se si elimina una variabile, si ottiene una approssimazione di \mathbf{W} . E' così che deve essere effettuata la semplificazione dei dati?

Varianza della 1^a componente

$$\hat{\sigma}^2(\mathbf{F}_1) = \frac{1}{n} \sum_{j=1}^n (F_{j1} - \bar{F}_1)^2 \quad \text{dove } \bar{F}_1 \text{ è la media del fattore } \mathbf{F}_1.$$

Inoltre

$$\bar{F}_1 = \frac{1}{n} \sum_{j=1}^n F_{j1} = \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^m q_{1k} \hat{\mathbf{x}}_{jk} = \frac{1}{n} \sum_{k=1}^m q_{1k} \sum_{j=1}^n \hat{\mathbf{x}}_{jk} = 0; \quad \text{dato che } \bar{\hat{\mathbf{x}}}_k = 0$$

e quindi

$$\hat{\sigma}^2(\mathbf{F}_1) = \frac{1}{n} \sum_{j=1}^n (F_{j1})^2 = \frac{1}{n} \|\mathbf{F}_1\|^2$$

cioè la norma euclidea al quadrato del nuovo fattore \mathbf{F}_1 (divisa per n). Ne consegue che

$$\hat{\sigma}^2(\mathbf{F}_1) = \frac{1}{n} \mathbf{F}_1' \mathbf{F}_1 = \frac{1}{n} \mathbf{q}_1' \hat{\mathbf{x}}' \hat{\mathbf{x}} \mathbf{q}_1 = \mathbf{q}_1' \left(\frac{1}{\sqrt{n}} \hat{\mathbf{x}} \right)' \left(\frac{1}{\sqrt{n}} \hat{\mathbf{x}} \right) \mathbf{q}_1 = \mathbf{q}_1' \mathbf{W} \mathbf{q}_1$$

ovvero una forma quadratica basata sulla matrice di varianze-covarianze \mathbf{W} .

N.B. Il massimo non vincolato della varianza della componente è infinito.

Varianza della prima componente/2

La massimizzazione rispetto a tutti i possibili vettori è eccessiva; occorre almeno tenere conto che le matrici di rotazione contengono vettori normalizzati. Questo equivale ad imporre il vincolo

$$\|q_1\| = \sqrt{\sum_{j=1}^m q_{j1}^2} = 1 \Rightarrow -1 \leq q_{j1} \leq 1, j = 1, 2, \dots, m$$

A questo punto, la massimizzazione della varianza si configura come un massimo vincolato e perciò ricorriamo ai moltiplicatori di Lagrange

$$\begin{aligned} & \max \left\{ \hat{\sigma}^2(F_1) - \lambda(q_1^t q_1 - 1) \right\} \\ & \left\{ \begin{aligned} \frac{\partial}{\partial q_1} [q_1^t W q_1 - \lambda(q_1^t q_1 - 1)] &= 0 \Rightarrow 2Wq_1 - \lambda 2I_m q_1 = 0 \\ \frac{d}{d\lambda} [q_1^t W q_1 - \lambda(q_1^t q_1 - 1)] &= 0 \Rightarrow q_1^t q_1 - 1 = 0 \end{aligned} \right. \end{aligned}$$

Variabilità spiegata dalla 1ª componente

Supponiamo di misurare la variabilità complessiva presente nella matrice dei dati con la traccia di W (che, come ricordiamo, contiene le varianze originarie)

$$\text{Variabilità complessiva} = Tr(W) = \sum_{i=1}^m \hat{\sigma}^2(X_i) \quad \text{con} \quad \hat{\sigma}^2(X_i) = \frac{1}{n} \sum_{h=1}^n (x_{hi} - \bar{x}_i)^2$$

Quanta parte della variabilità totale è spiegata dalla 1ª componente? Ricordiamo la relazione che sussiste tra autovalori e traccia.

$$Tr(W) = \sum_{i=1}^m \lambda_i \quad \text{con} \quad \lambda_i = \text{autovalore di } W$$

Inoltre $\hat{\sigma}^2(F_1) = q_1^t W q_1 = \lambda_1$

La quota di variabilità spiegata dalla componente F_1 è dunque

$$\alpha_1 = \frac{\lambda_1}{\sum_{i=1}^m \lambda_i}$$

che, come rapporto di composizione, varia tra zero: nessuna variabilità, spiegata dalla F_1 , ad uno cioè la matrice dei dati ha rango uno (un solo autovalore non nullo)

Varianza della prima componente/3

Dalla prima equazione si ottiene $Wq_1 = \lambda q_1$

cioè λ deve essere un autovalore della matrice di varianze-covarianze e q_1 il corrispondente autovettore (normalizzato, in modo da soddisfare anche il secondo vincolo sulla norma unitaria.

$$q_1^t q_1 = 1$$

La matrice di varianze-covarianze W è, per costruzione, simmetrica e, almeno, positiva semidefinita per cui prevede m autovalori reali non negativi. Quale è quello che risolve il problema?

Poiché $q_1^t W q_1 = \lambda q_1^t q_1 = \lambda$ (dato che $q_1^t q_1 = 1$)

allora λ non può che essere l'autovalore massimo di W e q_1 è l'autovettore normalizzato ad esso collegato.

Quindi, la prima componente è $F_1 = \hat{X} q_1$

Dove F_1 è il vettore su cui sono proiettati gli "n" punti rispetto al quale la nube dei punti originali esprime la massima dispersione.

Esempi

1) Colonne di costanti

$$X = \begin{bmatrix} 1 & 12 & -3 \\ 1 & 12 & -3 \\ 1 & 12 & -3 \\ 1 & 12 & -3 \\ 1 & 12 & -3 \end{bmatrix} \Rightarrow \hat{X} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow W = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{aligned} \lambda_{\max} &= 0 \\ \alpha_1 &= 0 \end{aligned}$$

La quota di variabilità spiegata dalla 1ª componente è $0/(0+0+0)=\text{indeterminata}$ (0)

2) Colonne di costanti tranne una

$$X = \begin{bmatrix} 1 & 1 & -3 \\ 4 & 1 & -3 \\ 2 & 1 & -3 \\ 1 & 1 & -3 \end{bmatrix} \Rightarrow \hat{X} = \frac{1}{2} \begin{bmatrix} -1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \Rightarrow W = \begin{bmatrix} \frac{6}{4} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{aligned} \lambda_{\max} &= 6/4 \\ \alpha_1 &= 1 \end{aligned}$$

La quota di variabilità spiegata dalla 1ª componente è $(6/4)/(6/4+0+0)=1$

Lettura dei pesi della 1^a componente

$$F_1 = q_{11}\hat{X}_1 + q_{12}\hat{X}_2 + \dots + q_{1m}\hat{X}_m = \hat{X}q_1$$

Premoltiplichiamo per la matrice centrata trasposta

$$\hat{X}^t F_1 = (\hat{X}^t \hat{X})q_1 \Rightarrow nWq_1 = n\lambda_1 q_1 \Rightarrow n \begin{bmatrix} \lambda_1 q_{11} \\ \lambda_1 q_{12} \\ \vdots \\ \lambda_1 q_{1m} \end{bmatrix} = n \begin{bmatrix} \bar{x}_1^t F_1 \\ \bar{x}_2^t F_1 \\ \vdots \\ \bar{x}_m^t F_1 \end{bmatrix}$$

Da cui consegue l'identità $n\lambda_j q_{1j} = n\hat{X}_j^t F_1 \Rightarrow q_{1j} = \frac{\hat{X}_j^t F_1}{n\lambda_j} \quad j=1,2,\dots,m$

Richiamiamo l'interpretazione del prodotto scalare e scriviamo

$$q_{1j} = \frac{\hat{\sigma}(\hat{X}_j)}{\sqrt{\lambda_j}} \cdot \frac{\hat{X}_j^t F_1}{\hat{\sigma}(\hat{X}_j) \cdot \sqrt{\lambda_j}} = \frac{\hat{\sigma}(\hat{X}_j)}{\sqrt{\lambda_j}} r(\hat{X}_j, F_1)$$

Che misura l'angolo formato dalla F_1 con la j -esima variabile.

La quantità $r(\hat{X}_j, F_1)$ è il coefficiente di correlazione tra le variabile originaria j -esima e la 1^a componente. Tale formula ci aiuterà ad interpretare i pesi q_{1j} .

Determinazione della 2^a componente/2

Tuttavia, poiché

$$Wq_1 = \lambda_1 q_1, \quad \text{si ha anche } q_2^t Wq_1 = \lambda_1 q_2^t q_1 = 0$$

Pertanto, si ottengono nuove variabili incorrelate purché i pesi delle componenti siano ortogonali

La nuova funzione da massimizzare rispetto ai pesi della componente è

$$q_2^t (W - \lambda_1 q_1 q_1^t) q_2 - \lambda_2 q_2^t q_2$$

Che equivale al problema di massimo vincolato

$$q_2^t Wq_2 - \lambda_1 q_1 q_1^t q_2 = q_2^t Wq_2 - \lambda_2 q_2^t q_2$$

Infatti, il vincolo di ortogonalità è di fatto incorporato nella funzione da massimizzare.

Determinazione della 2^a componente

La variabilità spiegata dalla 1^a componente può risultare insoddisfacente ed occorre aggiungere una seconda componente.

$$F_2 = \hat{X}q_2$$

Non si può, a questo fine, massimizzare la varianza totale dato che non si farebbe altro che ricalcolare esattamente la 1^a componente.

Occorre quindi massimizzare la variabilità residua dopo che la prima componente sia già stata determinata.

$$q_2^t [W - \lambda_1 q_1 q_1^t] q_2$$

Poiché si parte da un sistema ortogonale è opportuno (ma non necessario) ottenere un nuovo sistema ortogonale e cioè

$$F_2^t F_1 = 0 \Rightarrow q_2^t \hat{X}^t \hat{X} q_1 = q_2^t Wq_1 = 0 \Rightarrow q_2^t Wq_1 = 0$$

Determinazione della 2^a componente/3

Anche per il secondo vettore di pesi si impone la condizione di norma unitaria:

$$q_2^t q_2 = 1$$

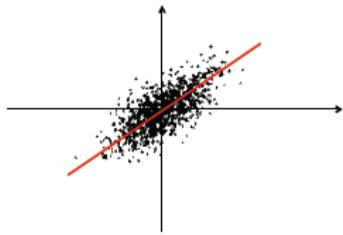
A questo punto il calcolo della seconda componente può procedere come per la prima

$$Wq_2 - \lambda_2 q_2 = 0 \quad \text{con } q_2^t q_2 = 1$$

e quindi λ deve essere un altro autovalore di W e q_2 è il corrispondente autovettore.

Poiché si tratta ancora di un problema di massimizzazione, il valore di λ non può che essere il secondo autovalore di W in ordine decrescente di grandezza cioè $\lambda = \lambda_2$

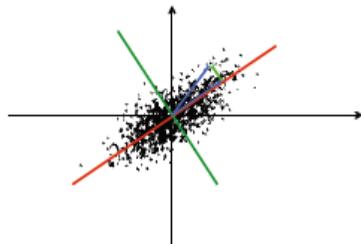
Simple illustration of PCA



First principal component of a two-dimensional data set.

illustrazione

Simple illustration of PCA



Second principal component of a two-dimensional data set.

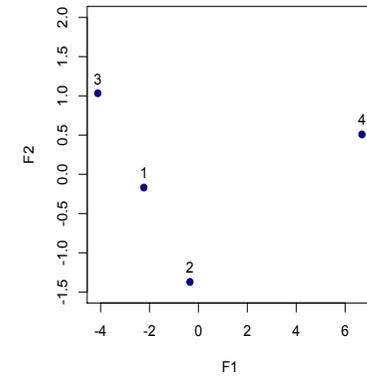
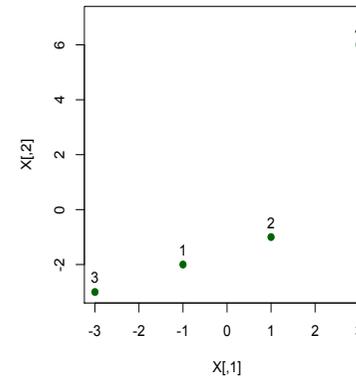
Esempio

$$X = \begin{bmatrix} 2 & 2 \\ 4 & 3 \\ 0 & 1 \\ 6 & 10 \end{bmatrix}; \hat{X} = \begin{bmatrix} -1 & -2 \\ 1 & -1 \\ -3 & -3 \\ 3 & 6 \end{bmatrix}; W = \begin{bmatrix} 5 & 7 \\ 7 & 12.5 \end{bmatrix}; \lambda = \begin{bmatrix} 16.6912 & 0.8088 \end{bmatrix}$$

$$\alpha = \begin{bmatrix} 95.4\% & 4.6\% \end{bmatrix}$$

$$q_1 = \begin{bmatrix} 0.5137 & 0.8580 \end{bmatrix}; q_2 = \begin{bmatrix} -0.8580 & 0.5137 \end{bmatrix}$$

Vecchi e nuovi assi



Ulteriori componenti

In modo analogo si possono derivare le successive componenti

$$F_j = \hat{X}q_j, \quad j = 1, 2, \dots, p \leq m$$

Dove q_j è l'autovettore normalizzato (vettore dei pesi) associato all'autovalore j -esimo in ordine decrescente di grandezza.

Inoltre

$$1) \text{var}(F_j) = \text{var}(q_j \hat{X}) = \lambda_j, \quad j = 1, 2, \dots, p$$

$$2) \text{cov}(F_i, F_j) = 0 \quad \forall i \neq j$$

N.B. La seconda condizione richiede che gli autovalori di W siano tutti distinti.

The Painter's Data of de Piles

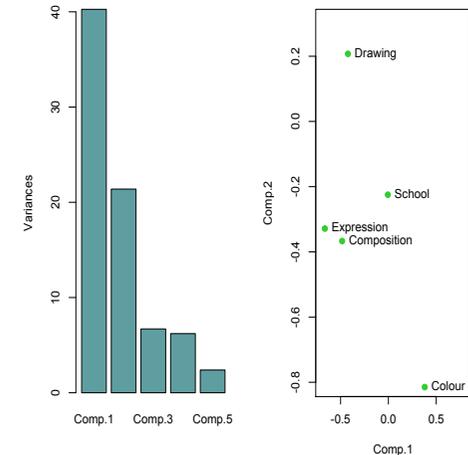
> painters

	Composition	Drawing	Colour	Expression	School
Da Udine	10	8	16	3	A
Da Vinci	15	16	4	14	A
Del Piombo	8	13	16	7	A
Del Sarto	12	16	9	6	A
Fr. Penni	0	15	8	0	A
Giulio Romano	15	16	4	14	A
Michelangelo	8	17	4	8	A
Perino del Vaga	15	10	7	6	A
Perugino	4	12	10	4	A
Raffaël	17	18	12	18	A
F. Zucarno	10	13	8	8	B
Fr. Salviata	13	15	8	8	B
Parisiqano	18	15	6	6	B
Prinaticcio	15	14	7	10	B
T. Zucarno	13	14	10	9	B
Volterra	12	15	5	8	B
Bonocci	14	15	6	10	C
Corona	15	14	12	6	C
Giogione	18	18	6	2	C
L. Jordaens	13	12	9	6	C
Testa	11	15	8	6	C
Varius	15	15	12	13	C
Barsoo	6	9	17	0	D
Bellini	4	5	14	0	D
Giorgione	8	9	18	4	D
Murillo	6	9	15	4	D
Palma Giovane	12	9	14	0	D
Palma Vecchio	5	6	16	0	D
Fondone	8	14	17	5	D
Tintoretto	15	14	10	4	D
Titian	15	15	18	6	D
Venese	15	19	10	3	D
Ribani	14	14	18	6	E
Caravaggio	6	5	16	8	E
Caraccio	13	13	15	12	E
Donnichino	15	17	9	17	E
Guercino	18	18	18	4	E
Luffranco	14	13	18	5	E
The Carracci	15	17	13	13	E
Durer	8	10	18	8	F
Holbein	9	10	16	18	F
Pourbus	4	15	5	6	F
Van Leyden	8	5	5	4	F
Diapenback	11	19	14	6	G
J. Jordaens	18	8	16	6	G
Otho Varius	13	14	18	18	G
Reinhardt	15	5	17	12	G
Rubens	18	13	17	17	G
Taniens	15	12	13	6	G
Van Dyck	15	19	17	13	G
Bourdon	18	8	9	4	H
La Bruin	16	16	8	16	H
Le Suer	15	15	4	15	H
Poussin	15	17	6	15	H

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	6.3453289	4.624592	2.58866562	2.4912964	1.54665234
Proportion of Variance	0.5232389	0.277932	0.08788506	0.0806571	0.03108688
Cumulative Proportion	0.5232389	0.801171	0.88825602	0.9689131	1.00000000

PCA.Paint



Punteggi fattoriali

Le componenti sono delle pseudo-variabili che nascono dalle quote parti che ogni variabile originaria trasferisce nella comune componente.

La componente non è osservabile, ma ogni unità riceve uno pseudo-valore ovvero un punteggio o uno score ad essa riferita.

I punteggi fattoriali sono inglobati nella matrice F di dimensione (n x p) per cui i punteggi sono legati alle variabili originarie attraverso i pesi Q.

Tali pesi sono proporzionali alle correlazioni

$$E(\hat{X}'F) = E(\hat{X}'\hat{X}Q'_p) = E(\hat{X}'\hat{X})Q'_p = RQ'_p$$

dove R è la matrice di correlazione tra le variabili originarie.

La percentuale di variabilità totale riprodotta da complesso di p fattori è misurata da

$$\alpha_p = \frac{\sum_{i=1}^p \lambda_i}{Tr(W)} \cdot 100$$

Lettura della matrice di rotazione o dei pesi

La matrice di rotazione è una matrice ortogonale che risolve l'equazione

$$F = \hat{X}Q^t \quad \text{con} \quad Q^tQ = I_p$$

Le nuove variabili F sono combinazioni lineari delle vecchie variabili X già presenti nella matrice dei dati.

Anche le vecchie variabili possono essere espresse come combinazioni lineari delle nuove variabili.

Partiamo dalle componenti espresse come combinazioni lineari delle variabili originarie

$$F_i = \sum_{j=1}^m q_{ij} X_j, \quad i = 1, 2, \dots, p$$

Il prodotto matriciale tra vecchie e nuove variabili è

$$FQ = \left(\hat{X}Q^t \right) Q = \hat{X} \quad \text{se } p=m$$

Dieta in alcuni Paesi europei. Contiene 9 variabili e 25 unità.

Esempio

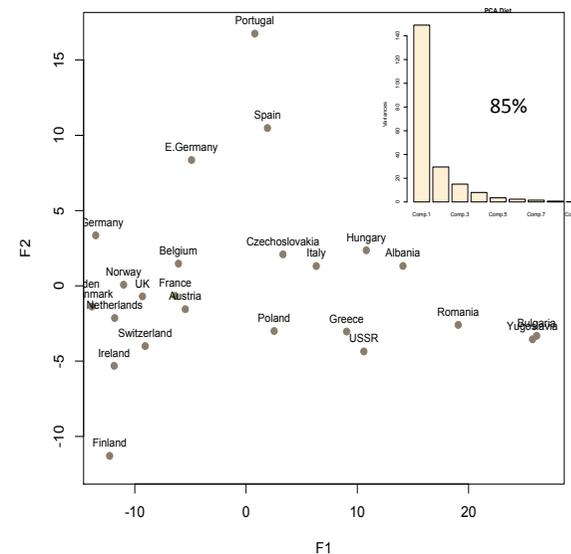
Pesi

	Comp. 1	Comp. 2
Red. Meat	-0.1587	-0.1327
White. Meat	-0.1295	-0.0434
Eggs	-0.0673	-0.0209
Milk	-0.4254	-0.8309
Fish	-0.1270	0.2923
Cereal	0.8609	-0.4062
Starch	-0.0669	0.0760
Nuts	0.1139	0.0701
Fr. Veg.	0.0282	0.1692

Componenti

	F1	F2
Albania	14.10	1.32
Austria	-5.46	-1.55
Belgium	-6.08	1.48
Bulgaria	26.12	-3.32
Czechoslovakia	3.32	2.09
Denmark	-13.86	-1.37
E. Germany	-4.90	8.36
Finland	-12.26	-11.29
France	-6.35	-0.67
Greece	9.04	-3.03
Hungary	10.80	2.36
Ireland	-11.86	-5.31
Italy	6.31	1.31
Netherlands	-11.81	-2.13
Norway	-11.01	0.08
Poland	2.53	-3.00
Portugal	0.78	16.75
Romania	19.07	-2.59
Spain	1.92	10.48
Sweden	-14.84	-0.73
Switzerland	-9.07	-4.00
UK	-9.31	-0.70
USSR	10.59	-4.35
W. Germany	-13.51	3.35
Yugoslavia	25.74	-3.55

Le unità nel piano nelle prime 2 comp.



Lettura della matrice di rotazione o dei pesi/2

Leggiamo la relazione da destra verso sinistra

$$X_j = \sum_{i=1}^p F_i q_{ij} = F_1 q_{1j} + F_2 q_{2j} + \dots + F_p q_{pj}$$

E riportiamo la relazione che lega la componente i-esima alle variabili originarie

$$F_i = q_{i1} \hat{X}_1 + q_{i2} \hat{X}_2 + \dots + q_{im} \hat{X}_m = \hat{X} q_i$$

I pesi compaiono in entrambe le espressioni ad indicare il legame tra la variabile j-esima e la componente i-esima.

I pesi non sono gli stessi: righe nel primo caso e colonne nel secondo

Poiché la matrice Q non è simmetrica il peso della variabile j sulla componente i è diverso dal peso della componente j sulla variabile i.

Standardizzazione delle componenti

Per ragioni di chiarezza le componenti principali sono calcolate in modo da rendere la loro norma (e quindi la loro varianza) pari ad uno.

A questo fine si utilizza la trasformazione

$$F_i^* = \frac{1}{\sqrt{\lambda_i}} \hat{X}q_i \quad q_i = (q_{i1}, q_{i2}, \dots, q_{im}); \quad i = 1, 2, \dots, p$$

E' agevole controllare che ora

$$\sigma^2(F_i^*) = \frac{1}{n} F_i^{*t} F_i^* = \frac{1}{\lambda_i} q_i^t \left(\frac{1}{\sqrt{n}} \hat{X}^t \hat{X} \frac{1}{\sqrt{n}} \right) q_i = \frac{1}{\lambda_i} q_i^t W q_i = \frac{\lambda_i}{\lambda_i} = 1, \quad i = 1, 2, \dots, p$$

La normalizzazione non sarebbe del tutto necessaria. In fondo le variabili originarie hanno una variabilità diversa e non si comprende perché le nuove dovrebbero avere varianza uguale e pari ad uno.

Una ragione è che così facendo i fattori sono standardizzati (media zero e scarto quadrato medio uno) facilitando la lettura degli score (cioè gli elementi di F).

Esercizio

Table 1. Correlation Matrix for Correlated Data

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.00	.97	.90	.04	-.14	-.05	-.18	.42	.41	.51
X2		1.00	.90	.28	.07	.20	.05	.46	.42	.60
X3			1.00	.17	.01	.10	-.01	.42	.37	.53
X4				1.00	.94	.95	.95	.49	.38	.49
X5					1.00	.93	.98	.57	.48	.52
X6						1.00	.97	.47	.43	.49
X7							1.00	.45	.34	.41
X8								1.00	.94	.96
X9									1.00	.89
X10										1.00

In che modo la ACP riflette la ipotesi di correlazione e incorrelazione di queste due matrici?

Table 2. Correlation Matrix for Noncorrelated Data

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.00	-.11	.20	-.03	.05	.37	-.21	.49	.43	.54
X2		1.00	.33	.68	.06	.39	.42	.16	-.01	.57
X3			1.00	.30	.46	.34	.51	.28	.17	.47
X4				1.00	.25	.41	.51	.40	.06	.44
X5					1.00	.36	.31	.66	.46	.20
X6						1.00	-.21	.36	.68	.53
X7							1.00	.15	-.28	.22
X8								1.00	.39	.51
X9									1.00	.07
X10										1.00

Matrice dei pesi o loadings

Le nuove variabili possono essere riscalate rispetto alla varianza delle vecchie

$$F^+ = \hat{X}P, \quad P = Q^t D^{0.5} \Rightarrow \text{dove } D^{0.5} = \text{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_m}\}$$

Che può essere vista come una forma alternativa di standardizzazione

$$\begin{aligned} \sigma^2(F_i^+) &= F_i^{+t} F_i^+ = \sqrt{\lambda_i} (\hat{X}q_i)^t \hat{X}q_i \sqrt{\lambda_i} \\ &= \lambda_i q_i^t W q_i = \lambda_i^2 \Rightarrow \sigma(F_i^+) = \lambda_i \end{aligned}$$

L'autovalore è lo scarto quadratico medio e non più la varianza.

Quindi le componenti ottenute con i pesi riscalati hanno media nulla, ma scarto quadratico medio distinto e coincidente con il loro autovalore. Tale scelta conserva la naturale diversità tra le variabilità delle componenti.

Notiamo inoltre che

$$P_{ji} = \sqrt{\lambda_i} \cdot q_{ij} = \sigma_j r(\hat{X}_j, F_i)$$

Per cui i pesi P_{ij} sono proporzionali allo scarto ed alla correlazione tra la variabile ed il fattore.

Letture della matrice dei pesi

La somma dei quadrati di righe e colonne hanno un preciso significato.

Se ci si ferma a $p < m$ fattori abbiamo

$$\begin{aligned} \text{Riga: } \sum_{j=1}^m P_{ij}^2 &= \sum_{j=1}^m \lambda_i q_{ji}^2 = \sum_{j=1}^m \sigma_i^2 r^2(\hat{X}_j, F_i) \\ &= \sigma_i^2 \sum_{j=1}^m r^2(\hat{X}_j, F_i) = \sigma_i^2 \end{aligned}$$

$$P = \begin{matrix} F_1^+ & F_2^+ & \dots & F_i^+ & \dots & F_p^+ \\ \hat{X}_1 & P_{11} & P_{12} & \dots & P_{1i} & \dots & P_{1p} \\ \hat{X}_2 & P_{21} & P_{22} & \dots & P_{2i} & \dots & P_{2p} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \hat{X}_j & P_{j1} & P_{j2} & & P_{ji} & & P_{jp} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \hat{X}_m & P_{m1} & P_{m2} & & P_{mi} & & P_{mp} \end{matrix}$$

Da notare che ora la somma dei quadrati di colonna è diversa e l'entità dei pesi deve essere rapportata a tale differenza.

dato che ogni correlazione è una quota parte di correlazione della j -esima variabile con l' i -esimo fattore che in totale somma ad uno.

$$\text{Colonna: } \sum_{i=1}^m P_{ij}^2 = \sum_{i=1}^m \lambda_i \cdot q_{ij}^2 = \lambda_i \sum_{i=1}^m q_{ij}^2 = \lambda_i$$

dato che i pesi hanno norma unitaria.

Comunalità

L'idea è che la variabile X_j sia spiegata -chi più chi meno- da ciascuna delle p componenti.

Poiché di fattori se ne considerano solo p , quanta parte della singola X_j è spiegata da essi?

$$\text{Comunalità } \beta_i = \sum_{j=1}^p P_{ij}^2 \quad i = 1, 2, \dots, m$$

È chiaro che se la comunalità è piccola cioè lontana dall'unità vuol dire che i p fattori già inclusi non tengono conto abbastanza di X_j ed uno o più altri fattori debbono essere aggiunti ovvero X_j deve essere esclusa perché poco legata alle altre variabili.

il segno dei pesi

Il segno dei pesi è poco rilevante in quanto i vincoli posti sui coefficienti nella matrice dei pesi P riguardano la somma dei quadrati e questa rimane invariata rispetto al cambiamento di segno.

$$P = Q^t D^{0.5} \Rightarrow -P = -Q^t D^{0.5} \text{ e } (-P)^t (-P) = P^t P = D$$

Alcuni packages invertono il segno di tutti i pesi qualora prevalga quello negativo.

Questo dovrebbe facilitare l'interpretazione delle componenti qualora le relazioni lineari concordanti tra le variabili risultassero maggioritarie.

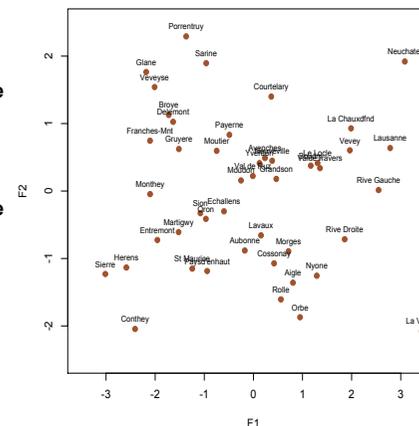
Data set sui cantoni svizzeri

Loadings:	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Fertility	-0.457	0.322		0.536	0.383	-0.473
Agriculture	-0.424	-0.412		-0.643	0.375	-0.309
Examination	0.510				0.814	
Education	0.454		-0.532			-0.681
Catholic	-0.350		-0.807			0.402
Infant.Mortality		0.811		-0.527		
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.167	0.167	0.167	0.167	0.167	0.167
Cumulative Var	0.167	0.333	0.500	0.667	0.833	1.000

Esempio

	Comunalità
Fertility	0.3125
Agriculture	0.3493
Examination	0.2755
Education	0.2385
Catholic	0.1439
Infant.Mortality	0.0000

Le unità nel piano delle prime due CP



Anche il valore numerico dei pesi è utile a comprendere la natura della componente.

Più grande è P_j maggiore è il legame lineare tra componenti ovvero maggiore è il contributo che la variabile fornisce alla costituzione del fattore.

Da notare che le prime due componenti non danno alcuna spiegazione sui valori osservati di mortalità infantile.

Ordinamento rispetto alla 1ª componente

Se la prima componente è un riassunto accettabile dei dati complessivi (diciamo con una variabilità spiegata superiore al 50%) conviene ordinare i punteggi fattoriali in senso crescente di grandezza

$$F_{(1)} \leq F_{(2)} \leq \dots \leq F_{(n)}$$

Secondo il grande M. Kendall, questo ordinamento è molto simile a quello in base al rango medio cioè la posizione definita ordinando le unità rispetto ad ogni variabile per poi sommare le singole posizioni in ogni variabile.

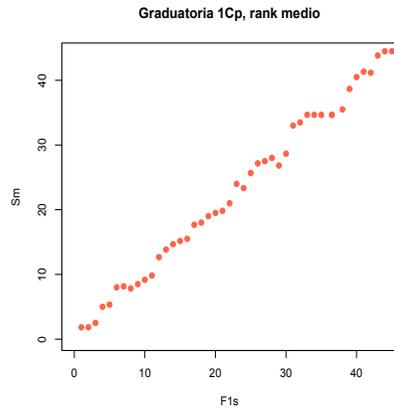
Per usare correttamente tale graduatoria occorrerà stabilire il verso-positivo o negativo - rispetto al fenomeno in esame.

In questo senso occorre guardare più all'entità numerica dei pesi ed ai valori originari.

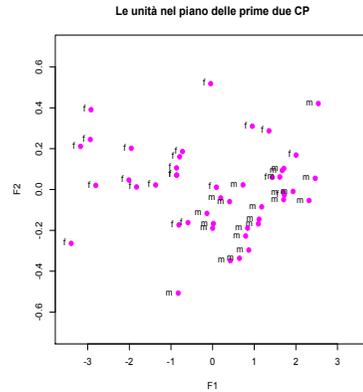
Jolicoer-Mosimann data set

Importance of components:
 Comp.1 Comp.2 Comp.3
 Standard deviation 1.71340 0.206988 0.1463505
 Proportion of Variance 0.97858 0.014281 0.0071395
 Cumulative Proportion 0.97858 0.992861 1.0000000

Loadings:
 Comp.1 Comp.2 Comp.3
 length -0.579 -0.325 0.748
 width -0.578 -0.483 -0.657
 height -0.575 0.813 0.000



Esempio



La congettura di Kendall non è smentita. Infatti il coefficiente di Spearman (che misura la correlazione tra i ranghi) è pari a 0.9981

PRINCIPAL COMPONENT ANALYSIS ON 88 INDIVIDUAL ITEMS WITH VARIMAX ROTATION OF 7 FACTORS*

Components ^a and variables	Factors						
	1	2	3	4	5	6	7
1 Give us a strong, decisive government	.67						
1 Give us a clear, consistent set of long-run policies	.66						
1 Bring respect to the nation	.63						
1 Develop a long-term global policy	.63						
1 Bring fresh, new ideas into office	.59						
1 Be able to say no to a big-spending Congress	.48						
1 Build up our military in every category		.66					
1 Reject SALT II		.50					
1 Reinstate the Draft		.50					
1 Increase our energy supply by building more nuclear plants		.47					
1 Reduce inflation by imposing a gasoline tax			-.64				
1 Fight for ERA			-.64				
2 Conservatives			.75				
2 Businessmen			.66				
2 Rich people			.49				
2 Independents			-.48				
2 Environmentalists			-.61				
2 Liberal Republicans			-.71				
2 Students			-.76				
3 Excluded				.57			
3 Relieved				.54			
3 Patriotic				.54			
3 Responsible				.53			
3 Confident				.53			
3 Optimistic				.50			
3 Satisfied				.49			
3 Hopeful				.47			
3 Involved				.46			
4 Trustworthy					.76		
4 Strong-minded					.73		
4 A man of high integrity					.68		
4 A man with a sense of purpose					.62		
4 Energetic					.61		
4 Stable					.61		
4 Compassionate					.53		
4 Articulate					.52		
5 Another U.S. Embassy is overtaken					.71		
5 All of the hostages in Iran are released					.67		
5 The Soviets invade another country					.63		
5 Economists predict a deep recession					.59		
5 The Afghanistan situation is rectified					.58		
5 We get to a point where we are on the brink of war					.58		
5 The inflation rate drops to an annual rate of 8%					.55		
5 Gasoline prices rise above \$2.00 a gallon					.51		
6 Was engaged in a political scandal						.81	
6 Evaded his taxes						.58	
6 He lied to the press						.52	
6 Was having romantic affairs with other women						.51	
6 Got caught drinking while driving						.49	
7 Of his media coverage							.52
7 Of his standing in the polls							.47
7 Of a bandwagon effect							.45

* Factor loadings of 0.45 and greater are presented.
^a Component 1: Issues and policies 5: Current events
 2: Social history 6: Personal events
 3: Emotional feelings 7: Epistemic issues
 4: Candidate image

Applicazione

Comportamento degli elettori in una votazione per le primarie. Indagine su 2000 rispondenti con ritorno di 655 questionari

Presenza di autovalori eguali

La simmetria della matrice di varianze-covarianze garantisce che gli autovalori da essa ricavati siano tutti reali e che le componenti siano pure reali e tra di loro ortogonali.

Non c'è però garanzia che gli autovalori siano distinti.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

Se λ_i è l'autovalore i -esimo in ordine decrescente di grandezza di W è possibile che $\lambda_{(i-1)} = \lambda_i$ oppure che $\lambda_{(i+1)} = \lambda_i$.

Gli autovettori associati ad autovalori multipli non possono essere determinati univocamente (ci sono uno o più gradi di libertà, a seconda della molteplicità dell'autovalore) e la varianza ad essi associata è la stessa. Si sa solo che sono ortogonali

E' chiaro che essi andranno inseriti o esclusi in blocco dalle componenti da trattenere per le analisi successive.

Presenza di autovalori eguali/2

Ad esempio, se gli indicatori dessero luogo ad una matrice di equicovarianza.

$$W = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} \text{ con } 0 < \rho \leq 1$$

all'autovalore massimo $\lambda_1 = \sigma^2[1+(m-1)\rho]$ è associato l'autovettore normalizzato: $q_1 = [m^{-0.5}, m^{-0.5}, \dots, m^{-0.5}]$ che spiega il $100[1+(m-1)\rho/m]$ di variabilità totale.

Gli altri autovalori sono tutti uguali tra di loro (infatti: $\lambda_i = \sigma^2(1-\rho)$ per $i = 2, 3, \dots, m$) e gli autovettori sono una qualsiasi delle $(m-1)$ soluzioni dell'equazione

$$\sigma^2 \rho (q_{12} + q_{22} + \dots + q_{m2}) = 0$$

Se volessimo aumentare il grado di copertura della variabilità spiegata non sapremmo quale componente scegliere e si rimarrebbe in posizione di stallo.

Presenza di autovalori eguali/3

La molteplicità degli autovalori è improbabile per dati incontrati nelle applicazioni reali.

Anzi, si pone spesso il problema contrario, e cioè che differenti rilevazioni dalla stessa popolazione producono stime molto diverse degli autovalori.

Se ad apparire uguali (almeno a livello di dati campionari) sono gli ultimi autovalori si parla di sfericità nelle ultime componenti che quindi non appaiono legate a particolari variabili o gruppi ristretti di variabili.

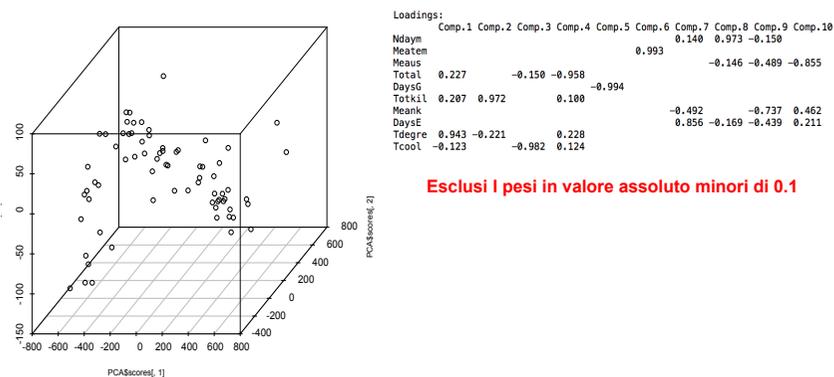
Piuttosto sono componenti indifferenziate irrilevanti nella spiegazione della variabilità dei dati che è invece assorbita dalle prime $p=m-k$ componenti.

Le ultime componenti tuttavia possono avere un ruolo importante ad esempio per riscontrare la presenza di valori anomali nei dati

Boston heating: n=71, m=10

Esempio

	Ndaym	Meatem	Meaus	Total	DaysG	Totkil	Meank	DaysE	Tdegre	Tcool
Jun-91	30	67	0.2	7	32	440	13.8	32	69	124
Jul-91	31	72	0.4	14	31	230	7.7	30	6	234
Aug-94	31	70	0.7	49	62	377	12.6	30	3	241
Aug-95	31	73	0.7	47	60	453	15.6	29	2	252
Sep-90	30	62	0.8	24	30	432	14.4	30	128	48



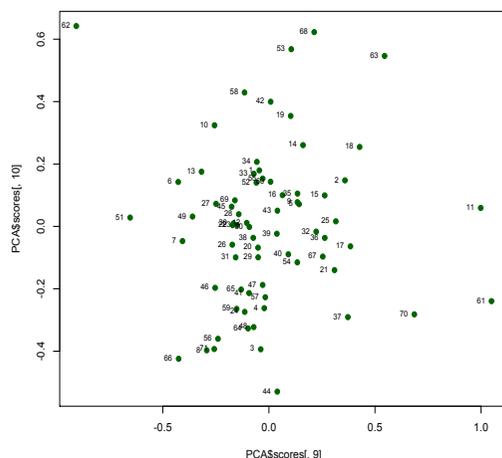
Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	417.1292957	151.8271113	51.25443697	40.074044845
Proportion of Variance	0.8640344	0.1144691	0.01304526	0.007974731
Cumulative Proportion	0.8640344	0.9785035	0.99154880	0.999523527

	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	8.9721020199	3.271741e+00	2.0062071184	7.509350e-01
Proportion of Variance	0.0003997404	5.315541e-05	0.0000199867	2.800234e-06
Cumulative Proportion	0.9999232675	9.999764e-01	0.9999964096	0.999992e-01

Esempio (continua)

L'unità 62 è remota nel piano delle ultime due componenti



I valori osservati in questa unità sono distanti dalle medie, anche tenuto conto dello scarto quadratico medio delle variabili. Occorre verificare i dati

	Ndaym	Meatem	Meaus	Total	DaysG	Totkil	Meank	DaysE	Tdegre	Tcool
Mean	30.423	46.352	5.311	166.859	33.831	507.606	16.737	30.282	570.676	37.887
Feb-97	28.000	36.000	9.700	283.000	29.000	853.000	30.500	28.000	806.000	0.000

Autovalori nulli o quasi nulli

Anche questa è una situazione rara per dati reali perché un autovalore nullo implica l'esistenza di una perfetta relazione lineare (o una perfetta sovrapposizione) tra due o più indicatori.

Tuttavia, a meno di errori nella definizione degli indicatori, è virtualmente impossibile trovare relazioni lineari esatte in dati reali.

Più realistico è il problema di individuare relazioni di quasi dipendenza.

In questi casi è facile trovare degli autovalori molto grandi (conseguenza delle alte collinearità fra gli indicatori) e contemporanea presenza di autovalori molto piccoli (la somma degli autovalori, come sappiamo è fissa).

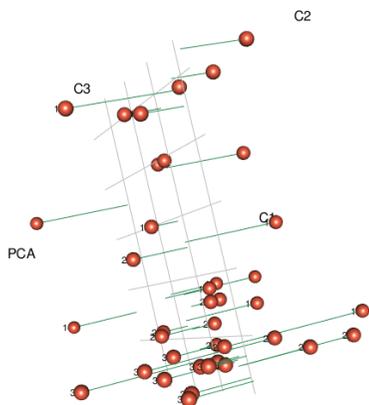
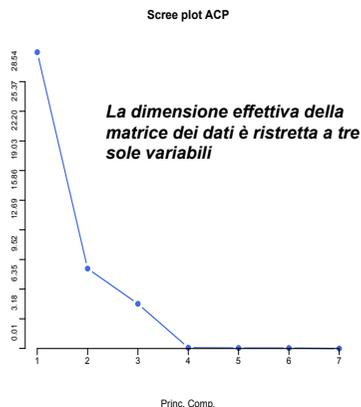
Uno o più autovalori quasi nulli significa che esistono variabili per i quali le unità sono poco differenziate e che perciò possono essere utilizzate, senza perdita sostanziale di informazione, solo le prime componenti.

Esempio

Composizione di sedimenti sabbiosi: 39 unità e 7 variabili

Relazioni lineari esatte o quasi non sono rare nelle composizioni

Importance of components:	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	31.6853	8.54649	4.78187	8.994e-02	5.859e-02	5.49e-02	3.040e-05
Proportion of Variance	0.9128	0.06641	0.02079	7.355e-06	3.121e-06	2.74e-06	8.403e-13
Cumulative Proportion	0.9128	0.97920	0.99999	1.000e+00	1.000e+00	1.00e+00	1.000e+00



Decathlon Data: n=33, m=10

Esempio

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
100m				0.168		-0.116	0.103	-0.233	0.944	
long				-0.136		0.274	0.908	0.280	0.057	
poid	0.156	-0.205	-0.177	0.887		-0.323	-0.053	0.053	0.056	
haut						0.110		0.050	-0.993	
400m		0.068	0.848	0.332		0.394		-0.083		
110mos			0.398	-0.186		-0.763	0.298	0.334	-0.122	-0.089
disq	0.118	0.352	-0.891	0.120	-0.222	0.067				
perc				-0.164	0.107	0.231	0.947	-0.076	-0.060	
jave	0.053	0.918	0.387		-0.071					
1500m	0.990	-0.095	0.089	-0.052						

I pesi della decima componente denotano il suo isolamento in quanto coincidenti con i risultati del salto in alto. Questi, peraltro, hanno peso trascurabile su tutte le altre componenti tranne che per un peso leggermente positivo sulla componente 6.

Componenti isolate

Se una delle variabili è del tutto incorrelata con le altre questo si rifletterà in una componente il cui autovettore avrà un solo elemento non nullo al posto corrispondente alla variabile isolata e con tutti gli altri elementi pari a zero.

Quindi ci sarà la perfetta identità $\text{indicatore} = \text{componente}$

Se le variabili originarie fossero incorrelate, allora le componenti non farebbero altro che riprodurre, in ordine di varianza, gli indicatori originari.

Le prime p componenti spiegherebbero ancora la stessa percentuale della variabilità totale: le prime otto componenti di dieci indicatori spiegano l'80% della variabilità totale, ma come si potrebbe decidere quali siano i due indicatori da trascurare?

E' noto che l'inclusione di variabili in eccesso induce la comparsa di componenti associate ad autovalori troppo grandi falsificando la loro reale importanza.

Qualche volta si configurano come componenti isolate

Matrici positive

Si parla di matrici simmetriche formate da elementi tutti positivi

$$w_{ij} > 0 \text{ per } i, j = 1, 2, \dots, m.$$

Le proprietà di tali matrici furono studiate da Perron nel 1907. In particolare

- ◆ L'autovalore massimo di W è compreso tra la minima e la massima somma di riga di riga o di colonna di W
- ◆ L'autovalore massimo è semplice e non ripetuto (molteplicità 1)
- ◆ All'autovalore massimo è associato un autovettore con pesi tutti dello stesso segno. Componente trasversale.
- ◆ Il secondo autovalore ha associato un autovettore che è di tipo bipolare dato che ha almeno un peso di segno diverso dagli altri per associare l'ortogonalità con il primo.

Matrici positive/2

Il segno di dei pesi della seconda, terza, quarta, etc. componente è a volte negativo ed a volte positivo con un numero di segni legati alla grandezza dei pesi.

Le componenti successive alla seconda sono dei contrasti in cui c'è la suddivisione per segno e valore paritaria dei pesi e che perciò esprime le contrapposizioni che rimangono tra due gruppi principali di variabili originarie, una volta che si sia eliminata l'influenza del fattore dimensionale e del fattore bipolare.

Si tratta pure di una presenza "classica" e non sorprendente visto che i pesi della prima componenti sono tutti dello stesso segno (ad esempio positivi) e la seconda deve essere a questa ortogonale.

Manca però una giustificazione teorica analoga a quella del fattore trasversale (si dimostra che nelle componenti diverse dalla prima deve essere presente almeno un cambiamento di segno nei pesi, ma il numero di tali inversioni può essere stabilito solo per matrici particolari).

Matrici bisimmetriche

La matrice controidentità è definita con degli uno solo nelle posizioni della controdiagonale.

La matrice quadrata A è bisimmetrica se è invariante nella pre e post moltiplicazione per una controidentità: $A=JAJ$

$$J = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 2 & 3 \\ 4 & 3 & 1 \end{bmatrix}; J = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}; JA = \begin{bmatrix} 4 & 3 & 1 \\ 3 & 2 & 3 \\ 1 & 3 & 4 \end{bmatrix} \Rightarrow JAJ = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 2 & 3 \\ 4 & 3 & 1 \end{bmatrix}$$

Il vettore a è simmetrico se $a=Ja$ ed è emisimmetrico se $Ja=-a$

$$a = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}; Ja = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}$$

La matrice bisimmetrica ha n/2 autovettori emisimmetrici e n/2 autovettori simmetrici.

$$a = \begin{bmatrix} 3 \\ 1 \\ -1 \\ -3 \end{bmatrix}; Ja = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ -1 \\ -3 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \\ 1 \\ 3 \end{bmatrix}$$

Esempio

Records per paese e per specialità della corsa: n=55, m=8

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
X100m	-0.020	-0.211	-0.029	-0.359	0.190	0.887	-0.052	-0.014
X200m	-0.042	-0.359	-0.018	-0.834	-0.049	-0.410	0.062	-0.004
X400m	-0.111	-0.828	-0.378	0.396	-0.012	-0.048	0.020	-0.009
X800m	-0.005	-0.023	0.005	-0.010	-0.011	-0.007	-0.261	0.965
X1500m	-0.014	-0.045	0.050	-0.016	-0.043	-0.067	-0.959	-0.262
X5k	-0.079	-0.130	0.336	0.019	-0.909	0.184	0.053	0.000
X10k	-0.181	-0.299	0.849	0.135	0.364	-0.068	0.046	0.005
Marathon	-0.973	0.181	-0.142	-0.028	0.007	0.004	-0.001	-0.001

La prima componente è trasversale, ma forse isolata poiché solo la maratona è realmente impattante.

La bipolarità nasce piuttosto dal contrasto tra le prime due componenti che da sole spiegano più del 99% della variabilità.

Il fatto rilevante è che queste strutture sono preannunciate cioè sappiamo che esistono prima di effettuare a ACP

Matrici bisimmetriche/2

Matrice di correlazione con struttura bisimmetrica

$$\begin{pmatrix} 1.000 & 0.968 & 0.953 & 0.927 & 0.898 \\ 0.968 & 1.000 & 0.989 & 0.968 & 0.923 \\ 0.953 & 0.989 & 1.000 & 0.991 & 0.951 \\ 0.927 & 0.968 & 0.991 & 1.000 & 0.966 \\ 0.898 & 0.923 & 0.951 & 0.966 & 1.000 \end{pmatrix}$$

Autovalori

4.814351 0.121732 0.048999 0.012379 0.002539

Autovettori

[,1]	[,2]	[,3]	[,4]	[,5]
-0.4408	0.6251	0.5646	-0.3099	-0.0074
-0.4504	0.3126	-0.3308	0.6774	-0.3621
-0.4538	0.0027	-0.3784	-0.0575	0.8048
-0.4508	-0.2869	-0.3689	-0.5984	-0.4694
-0.4401	-0.6551	0.5409	0.2894	0.0290

La composizione degli autovettori è piuttosto vicina a quella prevista. Ciò dipende dal fenomeno oppure dipende dallo schema che impone la ACP?

Dipendenza dall'unità di misura

La ACP si è finora basata sulla matrice di varianze-covarianze W.

Come è ben noto la covarianza e la varianza dipendono strettamente dall'unità di misura con cui sono state rilevate le variabili originarie: anni, lunghezze, pesi, distanze, tempi, conteggi vari.

Consideriamo una matrice 2x2

$$W = \begin{bmatrix} \sigma_1^2 & c \\ c & \sigma_2^2 \end{bmatrix} \Rightarrow \lambda = \frac{\sigma_1^2 + \sigma_2^2 + \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4c^2}}{2}$$

$$q_1 = \begin{bmatrix} \frac{c}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4c^2}} \\ \frac{c - \sigma_1^2}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4c^2}} \end{bmatrix}; q_2 = \begin{bmatrix} \frac{-c}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4c^2}} \\ \frac{\sigma_1^2 - c}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4c^2}} \end{bmatrix}$$

Se ora moltiplichiamo la variabile X, per una costante g>0, come si modificano i termini della ACP?

Dipendenza dall'unità di misura/2

$$W' = \begin{bmatrix} g^2\sigma_1^2 & gc \\ gc & \sigma_2^2 \end{bmatrix} \Rightarrow (\lambda_1^2 - \lambda_2^2)^2 = \left[(g^2\sigma_1^2 - \sigma_2^2)^2 + 4gc \right]$$

$$\frac{d(\lambda_1^2 - \lambda_2^2)^2}{dg} = \left[4g\sigma_1^2(g^2\sigma_1^2 - \sigma_2^2) + 4c \right]$$

Quando la covarianza è positiva (c>0) lo scarto tra i due autovalori aumenta, diminuisce o rimane ostante secondo che

$$g^2 > \left(\frac{\sigma_2}{\sigma_1}\right)^2, \quad g^2 < \left(\frac{\sigma_2}{\sigma_1}\right)^2, \quad g^2 = \left(\frac{\sigma_2}{\sigma_1}\right)^2$$

Ne consegue che possiamo alterare il risultato delle componenti principali modificando le unità di misura (ad esempio passando da Kg a grammi o viceversa).

ACP sulla matrice di correlazione

Pur esistendo diverse forme di standardizzazione i packages statistici offrono solo quella classica che si sostanzia in variabili trasformate che hanno tutte media zero e varianza uno

$$Z_{h,j} = \frac{X_{h,j} - \mu_x}{\sigma_x}, \quad j = 1, 2, \dots, m; \quad h = 1, 2, \dots, n$$

Quindi, l'originale diversità tra varianze delle variabili è stata -forzatamente- annullata.

A questa scelta fa da perfetto contraltare la standardizzazione dei punteggi fattoriali che è ora coerente proporre pure con media zero e varianza unitaria

$$F_{h,i}^* = \left(\frac{1}{\sqrt{\lambda_i}} \right) Zq_i, \quad i = 1, 2, \dots, p; \quad h = 1, 2, \dots, n$$

ACP sulla matrice di correlazione/2

La combinazione: variabili standardizzate/componenti standardizzati ha alcune utilità e produce delle semplificazioni.

Poiché ora $\text{Tr}(R)=m$, la percentuale di variabilità spiegata è data ora da

$$\alpha_i = \left(\frac{\lambda_i}{m} \right) 100, \quad i = 1, 2, \dots, p$$

Le comunalità sono pari a $\beta_j = \sum_{i=1}^p q_{i,j}^2$

La matrice dei pesi è ora uguale alla matrice di correlazione variabili/componenti con notevole semplificazione della lettura dei risultati

$$q_{i,j} = \text{cor}(X_i, F_j)$$

ACP sulla matrice di correlazione/3

E' importante sottolineare che autovalori ed autovettori della matrice di varianze-covarianze e della matrice di correlazione sono diversi.

Se invece di standardizzare sottraendo la media e dividendo per lo scarto quadratico si standardizza in un altro modo, ad esempio:

$$a) Z_{h,j} = \frac{X_{h,j} - \mu_x}{m_x}, \quad b) Z_{h,j} = \frac{X_{h,j} - \min(X_j)}{\max(X_j) - \min(X_j)}$$

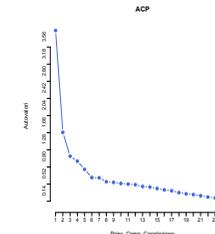
I risultati sono diversi (anche se non molto diversi).

Se le variabili originali sono tutte espresse nella stessa unità di misura la scelta della matrice di varianze-covarianze è accettabile.

Se le variabili hanno natura eterogenea allora standardizzare è un'opzione da non trascurare.

Esempio

Dati sulla complessione corporea:
n=507, m=25



La matrice di correlazione è positiva e quindi la struttura delle ACP prevede la componente trasversale, la bipolare ed i contrasti.

Loadings:

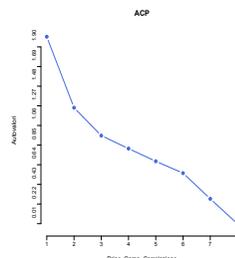
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Biac.dima	-0.18	0.18		0.19	0.31	-0.63	-0.24		-0.13	0.46
Biil.diam	-0.12	-0.30	-0.17	0.48	0.31	0.20		-0.64		0.10
Bitr.diam	-0.17	-0.26		0.36	0.10	-0.21	0.13	0.21	0.33	-0.48
Chest1	-0.20		-0.20	-0.21	0.13	0.26	-0.31		-0.43	-0.12
Chest2	-0.22				0.17	-0.27			0.25	
Elbow	-0.22	0.17		0.13			0.20	0.15		
Wrist	-0.21	0.17			-0.18			0.22	0.11	
Knee	-0.21		0.14	0.21	-0.24		0.34		-0.46	0.11
Ankle	-0.21	0.13		0.24	-0.22	0.34	0.17	0.15		0.40
Sgouldgir	-0.23	0.12		-0.18	0.13	-0.13				
Chestgir	-0.23			-0.24	0.14					
Waistgir	-0.23		-0.21	-0.20	0.12		-0.12			
Navel	-0.17	-0.32	-0.28	-0.14		0.19		0.15	0.13	0.37
Hipgir	-0.18	-0.38						0.29		
Thighgir	-0.12	-0.43	0.27	-0.21			0.17	0.15	-0.16	
Bicepgir	-0.23			-0.24			0.25	-0.16		
Forearmgir	-0.23	0.13		-0.14			0.16	-0.17		
Kneegir	-0.20	-0.20	0.22		-0.19		-0.23		-0.29	
Calfmagir	-0.20	-0.17	0.27		-0.29	-0.15	-0.23	-0.33		-0.19
Ankmagir	-0.20		0.18		-0.34		-0.47		0.46	0.17
Wrimagir	-0.23	0.16			-0.18					
Age			-0.73		-0.46	-0.29			-0.12	-0.13
Weight	-0.25				0.10					
Height	-0.19	0.16		0.36	0.22	0.24	-0.32	0.37	-0.13	-0.27
Sex	-0.19	0.34						-0.15		-0.19

Ogni entrata è una correlazione

Esempio

Heptathlon Data set n=18, m=8

Le unità di misura sono diverse perché ci sono tempi, lunghezze, pesi, conteggi. E' opportuno standardizzare.



Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
X100m	0.39		-0.33	-0.15	0.70	0.29	0.34	0.16
HighJump	-0.37	0.33	0.12	-0.39	0.53	-0.36	-0.35	-0.25
ShotPutt		-0.70	-0.13	0.45	0.36	-0.17	-0.27	-0.23
X200m	0.43		-0.16	-0.15	-0.12	-0.83	0.17	0.16
LongJump	-0.48		0.10	0.14		-0.17	0.81	-0.22
Javelin	-0.24	-0.36	-0.58	-0.60	-0.25	0.12		-0.21
X800m		-0.51	0.68	-0.46				0.19
Points	-0.49		-0.17					0.84

L'ottava componente è isolata e coincide con i Points.

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	2.0085	1.2498	0.9515	0.8129	0.67725	0.55054	0.275657	8.170e-03
Proportion of Variance	0.5043	0.1953	0.1132	0.0826	0.05733	0.03789	0.009498	8.343e-06
Cumulative Proportion	0.5043	0.6995	0.8127	0.8953	0.95261	0.99049	0.999992	1.000e+00

Scelta del numero di componenti

L'obiettivo della ACP è la determinazione di p variabili nuove che possano essere validamente sostituite alle m variabili originarie in tutte le analisi successive.

Peraltro, l'obiettivo si considera pienamente raggiunto se p è molto piccolo rispetto ad m e se il contenuto informativo dei due insiemi di indicatori, originali e nuovi, non differisce in maniera apprezzabile.

Si comprende perciò l'importanza di una accurata scelta del numero di componenti principali, ovvero di decidere quali siano gli autovalori "grandi" e quali quelli "piccoli".

I metodi disponibili per questa scelta sono diversi, nessuno dei quali privo di soggettività. Spesso, anzi, essi vanno usati in modo congiunto per non dar luogo a proposte arbitrarie:

Solo un valore di p su cui ci sia ampia convergenza nelle indicazioni che provengono da metodi diversi può essere accettato.

Percentuale cumulata di variabilità spiegata

Il criterio più immediato per la determinazione del numero di componenti è l'ammontare di variabilità spiegata dalle prime p componenti.

In genere l'ammontare di variabilità spiegata complessiva che si vuole raggiungere è legato al tipo di ricerca, ma raramente è inferiore al 60% o superiore al 90%.

Se così è p può essere dato dalla formula.

$$p = \text{Min}_{1 \leq i \leq m} \left\{ \alpha_i = 100 \left(\frac{\sum_{j=1}^i \lambda_j}{\sum_{i=1}^m \sigma_i^2} \right) \geq b \right\} \text{ con } 60\% \leq b \leq 90\%$$

Secondo questo criterio si continua ad estrarre componenti finché si arriva alla percentuale desiderata di variabilità spiegata.

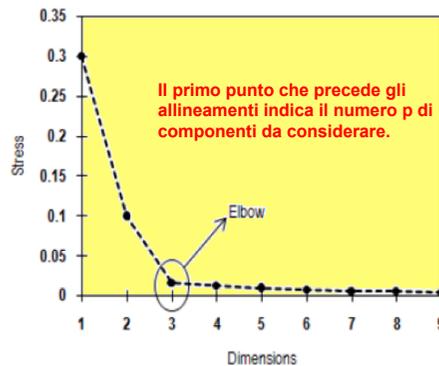
Poiché le componenti sono determinate in ordine di importanza, l'aggiunta di variabilità spiegata diminuisce aggiungendo una nuova componente, fino a diventare trascurabile.

Scree plot/2

Il grafico scree porta a valutare lo scarto tra autovalori per un numero di componenti crescenti: $S_k(X) - S_{k+1}(X)$

Il metodo consiste nel trovare il punto in cui l'inclinazione dei segmenti diventa quasi costante in modo che i segmenti sembrano far parte di una sola retta.

Questo vuol dire che i contributi che possono pervenire da queste nuove variabili sono disturbi che poco possono contribuire a migliorare la soluzione.



Il difetto del metodo è che in molti casi o non ci sono punti di svolta ben delineati oppure ve ne sono diversi.

In questi casi lo scree plot è di scarsa utilità.

Scree plot

A parità di condizioni la percentuale di variabilità spiegata diminuisce aumentando il numero di componenti, ma quella cumulata aumenta.

Quindi per migliorare la soluzione basterebbe aumentare il numero di componenti.

Tuttavia, il nostro obiettivo è ottenere la maggiore variabilità spiegata con il minor numero di componenti



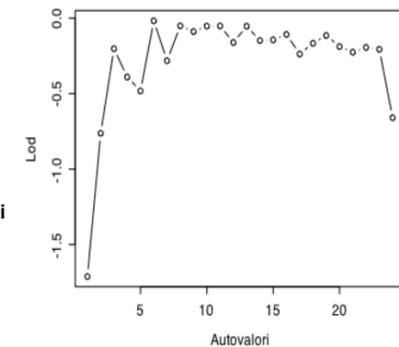
Grafico Lev

In alternativa si può usare un grafico dove le ordinate sono misurate in scala logaritmica (grafico LEV).

$$d_k = \log(\lambda_{k-1}) - \log(\lambda_k) = \log\left(\frac{\lambda_{k-1}}{\lambda_k}\right)$$

Il numero di componenti sarà indicato da un allineamento degli autovalori laddove una brusca indicazione che da lì in poi le componenti sono componenti erratiche che si possono ignorare.

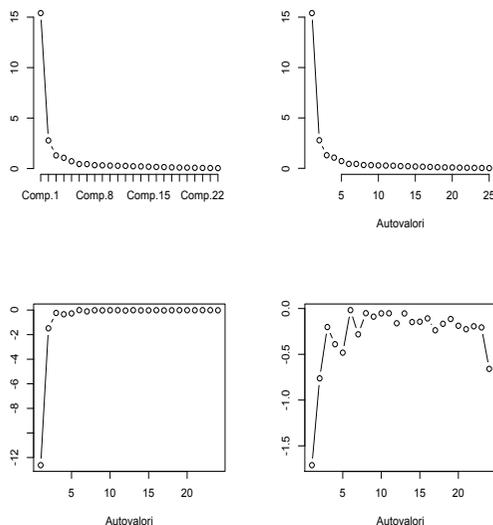
Da notare che nel grafico LEV si bada al rapporto tra autovalori successivi. Cioè si deve stabilizzare non più la differenza tra autovalori successivi, bensì il loro rapporto o variazione relativa



Body data set.

Esempio

Body.PCA



In base al grafico scree si può fermare a 6 componenti dato che da qui in poi l'incremento in variabilità spiegata è trascurabile.

Kaiser's rule

Poiché gli autovalori sono determinati in ordine decrescente di grandezza, la percentuale di variabilità spiegata dalla ultima componente inserita si riduce mano che ci si avvicina ad m .

Per scegliere p si potrebbe continuare a selezionare componenti finché l'ultima componente inserita non abbia un autovalore inferiore ad una soglia prefissata

$$p = \text{Max}_{1 \leq i \leq m} \{ \lambda_i \geq b \} \text{ con } b = \frac{\sum_{i=1}^m \sigma_i^2}{m}$$

Se le variabili originarie fossero standardizzate e quindi W coincidesse con la matrice di correlazione, vorrebbe dire trascurare tutte le componenti associate ad autovalori minori dell'unità.

E' stato però osservato che questo criterio sceglie meno componenti di quanto poi non si scopra necessario.

Sarebbe consigliabile orientarsi su valori di soglia più bassi. Ad esempio, con la matrice di correlazione, il valore di riferimento sarebbe l'unità, ma in molte occasioni è meglio usare 0.9 o addirittura 0.7.

Italian wine: $n=178, m=14$

Esempio

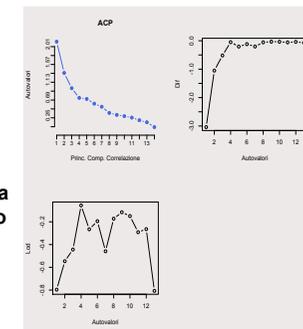
Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	2.3528595	1.5802140	1.2025283	0.96328492	0.93675104
Proportion of Variance	0.3954249	0.1783626	0.1032910	0.06627984	0.06267875
Cumulative Proportion	0.3954249	0.5737874	0.6770785	0.74335831	0.80603706
	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Standard deviation	0.8202307	0.74417670	0.59163686	0.54271718	0.51215828
Proportion of Variance	0.0480556	0.03955707	0.02500244	0.02103871	0.01873615
Cumulative Proportion	0.8540927	0.89364973	0.91865217	0.93969088	0.95842703
	Comp.11	Comp.12	Comp.13	Comp.14	
Standard deviation	0.47523512	0.41084877	0.359950245	0.240441928	
Proportion of Variance	0.01613203	0.01205691	0.009254584	0.004129451	
Cumulative Proportion	0.97455906	0.98661596	0.995870549	1.000000000	

Quattro componenti potrebbero bastare.

Esistono svariate tecniche e formalismi per decidere il numero di componenti.

La natura descrittiva e algebrica della ACP porta a preferire regole semplici e considerare il contesto applicativo per decidere se aumentare o diminuire il numero indicato.



Esempio

$N=250, m=20$

Component or Factor	Adjusted Eigenvalue	Unadjusted Eigenvalue	Estimated Bias
1	4.4965098	5.0115828	.51507294
2	3.6971236	4.1506432	.45351958
3	2.4540371	2.8268943	.37285721
4	1.4124735	1.6603032	.24782968
5	.86086266	1.0739862	.21312356
6	.50264634	.69026757	.18762124
7	.43161616	.57773755	.14612138
8	.4171857	.48482887	.06764317
9	.41289346	.46305685	.05016339
10	.38351215	.41568018	.03216803
11	.42688028	.38397364	-.04290664
12	.39338348	.34320168	-.05018181
13	.45403937	.31778023	-.13625914
14	.45767627	.29589049	-.16178578
15	.44264261	.25826351	-.1843791
16	.49499729	.24047616	-.25452113
17	.51687283	.23139579	-.28547704
18	.54348581	.20336817	-.34011763
19	.55823785	.18720304	-.3710348
20	.64292367	.18346663	-.45945704

La scelta dovrebbe cadere tra 4 o 5 componenti

Criterion: retain adjusted components > 1

The broken stick model

Una pertica di lunghezza unitaria è spezzata simultaneamente in p frammenti scegliendo a caso i ($p-1$) punti di rottura.

I frammenti sono disposti in ordine decrescente di lunghezza

Il valore atteso della lunghezza i -esima dei frammenti è

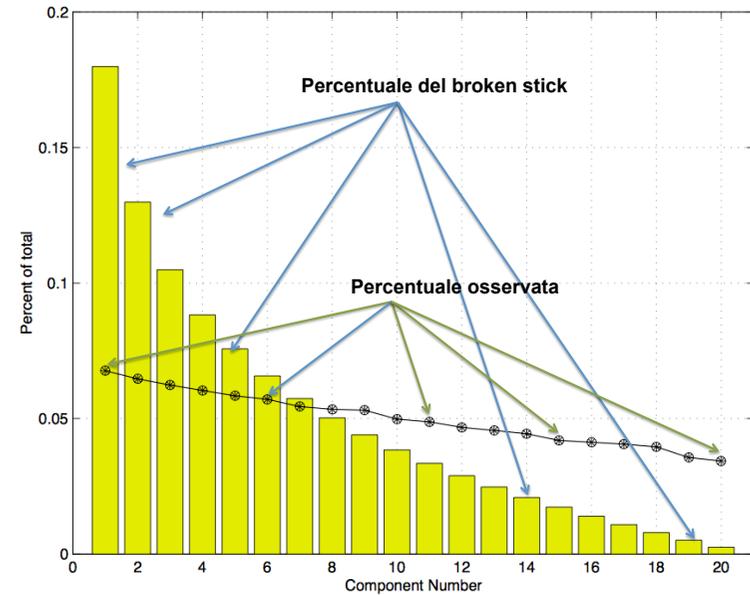
$$E(L_i) = \left(\frac{1}{m}\right) \sum_{j=i}^m \frac{1}{j}$$

	5	10	15	20
	45.67	29.29	22.12	17.99
	41.67	28.29	21.68	17.74
	36.67	27.18	21.20	17.48
	30.00	25.93	20.69	17.20
	20.00	24.50	20.13	16.90
		22.83	19.53	16.59
		20.83	18.86	16.26
		18.33	18.12	15.90
		15.00	17.29	15.52
		10.00	16.33	15.10
			15.22	14.64
			13.89	14.14
			12.22	13.59
			10.00	12.96
			6.67	12.25
				11.42
				10.42
				9.17
				7.50
				5.00

Se α_i è la percentuale di variabilità spiegata dall' i -esima componente il criterio diventa

$$p = \text{Max}_{i \leq m} \{ \alpha_i \geq b \} \text{ con } b = \left(\frac{100}{p}\right) \sum_{j=i}^p \frac{1}{j}$$

Esempio

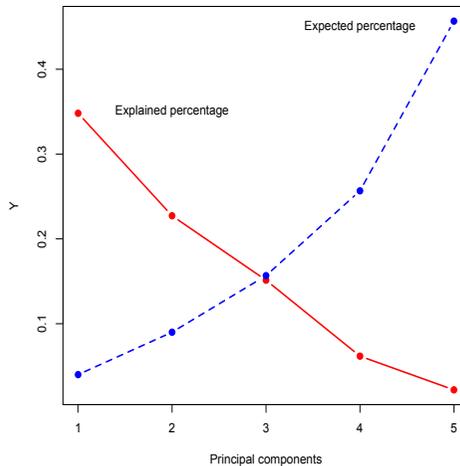


Altro esempio

Il numero di componenti è dato dalla posizione in cui le due spezzate si incrociano.

Nell'esempio, dalla 3ª in poi, l'aggiunta di variabilità spiegata è solo pari a quella che in media ci si aspetterebbe con delle componenti prive di struttura.

Quindi $p=2$ è una scelta ragionevole (forse anche $p=3$)



Rotazioni

La determinazione delle componenti avviene secondo i criteri della ortogonalità e della massimizzazione della variabilità complessiva che rimane ancora non spiegata dalle componenti principali già estratte.

Entrambi questi criteri possono essere tralasciati per ottenere configurazioni più facilmente interpretabili ovvero determinare una configurazione di pesi che preveda:



Pesi trascurabili per variabili poco rilevanti e pesi elevati per le variabili significative in ciascuna componente



Ogni variabile entri in maniera determinante in una sola componente e che nessun peso si collochi in grado intermedio.

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.7403724	1.1362825	0.7566080	0.3088664	0.110053797
Proportion of Variance	0.6057792	0.2582276	0.1144911	0.0190797	0.002422368
Cumulative Proportion	0.6057792	0.8640068	0.9784979	0.9975776	1.000000000

Rotazioni/2

Sono comunque fatti salvi il numero delle componenti e l'ammontare complessivo di variabilità da esse spiegata. Cambia soltanto la ripartizione all'interno delle componenti

Initial solution			Rotated solution		
f_1	f_2	f_3	f_1	f_2	f_3
x	x		x		
x	x	x	x		
x	x	x	x	x	
x		x		x	
x	x	x		x	
x	x		x	x	
x		x			x

Attenzione! L'effetto delle rotazioni è anche strettamente legato al numero di componenti e sarebbe consigliabile provare le rotazioni per diversi valori di p (numero di componenti).

Da notare che, ai fini della rotazione non è necessario coinvolgere le variabili originarie.

Ortogonalità della matrice di rotazione

Data l'ortogonalità della matrice dei pesi della ACP è vera la relazione

$$(Q^t Q)^{-1} (Q^t Q) = I$$

Ne consegue che

$$(Q^t Q)^{-1} (Q^t Q) T = (Q^t Q)^{-1} Q^t (QT) = (Q^t Q)^{-1} Q^t B$$

Che implica

$$T = (Q^t Q)^{-1} Q^t B$$

Il problema è che B non è nota e può essere determinata solo sulla base di principi soggettivi, analitici o grafici, legati al particolare metodo di rotazione che si sceglie (sono noti circa una ventina di metodi di rotazione).

Trasformazione della matrice di rotazione

Il problema della rotazione consiste nel moltiplicare la matrice dei pesi Q (che è già una matrice di rotazione) per un'altra matrice di rotazione che modifichi gli angoli scelti con la ACP.

$$QT = B$$

La matrice di rotazione T deve portare ad una nuova matrice di pesi di più consona agli obiettivi prefissati.

Se B fosse del tutto nota e Q fosse invertibile la determinazione di T sarebbe semplicemente

$$T = Q^{-1} B$$

Poiché il numero di componenti p è minore del numero di variabili m , la Q non può essere quadrata e T non può essere calcolata in questo modo.

No alle rotazioni oblique

L'ortonormalità di Q induce a ritenere che le componenti che collegano trasversalmente le variabili originarie non abbiano però legami di linearità vigenti tra di esse e quindi si chiede :

$$B^t B = I$$

Tale imposizione può apparire eccessiva: spesso le componenti sono riconducibili a macrofenomeni tra cui sono facilmente ipotizzabili dei rapporti di dipendenza, anche lineare.

Una rotazione di tipo obliquo che prevedesse perciò la possibilità di ottenere fattori correlati sarebbe più realistica.

Tuttavia, la ACP è una tecnica numerica che è forzosamente imposta ai dati e che ha finalità di semplificazione.

Le rotazioni oblique pur risultando più flessibili di quelle ortogonali, producono soluzioni più sofisticate e molto complicate da interpretare soprattutto per la difficoltà di definire e spiegare la cosiddetta "semplicità" della configurazione.

Quartimax

Va detto subito che i due principi con cui abbiamo articolato la nostra idea di “configurazione semplice” non sono facilmente conciliabili.

$$V_4 = \frac{1}{mp} \sum_{j=1}^p \sum_{i=1}^m (q_{ij}^2 - \bar{q}^2)^2$$

Ci si può muovere cercando una soluzione che punti a ridurre la complessità di riga in modo cioè che sia facilitata l'interpretazione, in termini di componenti principali, delle variabili.

In questo caso si può optare per il metodo Quartimax che porta a massimizzare la somma della potenza quarta dei pesi.

Il Quartimax riduce la presenza della variabile tra i fattori facendo sì che sia minimo il numero di fattori per cui la singola variabile ottiene pesi significativi.

Spesso si ottiene una soluzione in cui la prima componente è un fattore generale con pesi moderati o piccoli su tutte le variabili

Varimax

Si opta per la riduzione della complessità di colonna facilitando l'interpretazione delle componenti in termini delle variabili originarie.

$$V_1 = \sum_{j=1}^p \left(\sum_{i=1}^m q_{i,j}^4 - \frac{1}{m} \left(\sum_{i=1}^m q_{i,j}^2 \right)^2 \right)$$

Questo equivale a massimizzare la varianza del quadrato dei pesi in ciascuna colonna e orientarsi verso soluzioni in cui solo poche variabili hanno peso elevato sulla componente.

Dopo la rotazione varimax ciascuna delle variabili originarie tenderà ad essere associata con una o poche componenti ed ogni componente sarà espressione di un numero ridotto di variabili.

Inoltre, le componenti tendono a presentarsi come contrapposizioni tra variabili con pesi negativi e pesi positivi.

Esempio

Table 8.4. Audiometric Example. Rotation Matrices

Varimax	Quartimax
$\begin{bmatrix} .58 & .40 & .47 & .53 \\ -.52 & .70 & .39 & -.29 \\ .23 & .59 & -.77 & -.01 \\ .57 & -.00 & .18 & -.80 \end{bmatrix}$	$\begin{bmatrix} .82 & .38 & .43 & .04 \\ -.56 & .72 & .42 & .04 \\ .15 & .58 & -.79 & -.10 \\ -.00 & -.02 & .11 & -.99 \end{bmatrix}$

Table 8.5. Audiometric Example. Rotated Vectors

Frequency	Varimax	Quartimax
500L	.58 .13 .06 .71	.90 .11 .03 .22
1000L	.44 .10 .27 .79	.83 .09 .25 .36
2000L	.05 .22 .78 .45	.35 .22 .79 .28
4000L	.04 .89 .15 .20	.18 .89 .15 .11
500R	.91 .08 -.00 .23	.87 .05 -.07 -.35
1000R	.77 .10 .34 .31	.82 .08 .28 -.23
2000R	.17 .19 .93 -.00	.19 .19 .91 -.17
4000R	.11 .91 .19 -.02	.11 .90 .17 -.10

Altro esempio

Principal Component Analysis of Water Quality Data

Table 1a. Principal components from correlation matrix

	Latent Roots (Eigenvalues or Variances) Explained by Principal Components											
	1	2	3	4	5	6	7	8	9	10	11	12
	3.481	2.456	1.549	1.162	0.989	0.677	0.511	0.412	0.271	0.258	0.135	0.099
Percent of Total Variance Explained												
	1	2	3	4	5	6	7	8	9	10	11	12
	29.007	20.467	12.905	9.683	8.246	5.645	4.260	3.432	2.261	2.146	1.122	0.827
Component Loadings												
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Q	0.813	0.337	-0.262	-0.017	0.007	0.203	0.036	0.049	0.077	0.241	-0.168	-0.138
T	-0.734	0.539	-0.238	0.075	-0.035	0.018	0.084	0.182	0.023	-0.037	0.170	-0.191
pH	-0.088	0.061	-0.246	0.472	0.829	0.032	-0.041	-0.035	0.113	0.006	0.013	0.038
EC	-0.821	0.312	-0.067	0.079	-0.024	-0.203	-0.028	0.206	-0.160	0.304	-0.065	0.103
SS	0.644	0.575	-0.264	-0.003	-0.172	0.149	0.251	0.066	0.110	0.052	0.158	0.162
MA1	-0.463	-0.380	0.357	0.083	0.015	0.682	0.085	0.179	-0.025	0.016	-0.015	0.026
C1	-0.183	0.372	0.646	-0.349	0.301	-0.028	0.323	-0.274	-0.029	0.124	0.032	-0.032
NH ₃ -N	0.097	0.565	0.641	-0.037	-0.027	0.007	-0.450	0.095	0.213	0.032	0.035	0.015
NO ₃ -N	0.029	-0.378	-0.163	-0.775	0.333	-0.096	0.026	0.307	0.114	-0.015	0.003	0.014
DO	0.659	-0.623	0.113	0.067	0.095	-0.017	-0.161	0.023	-0.174	0.220	0.208	-0.048
Pv	0.524	0.675	0.061	-0.120	0.244	0.096	-0.091	0.131	-0.349	-0.184	-0.030	0.009
BOD ₅	0.467	-0.176	0.539	0.421	-0.007	-0.295	0.298	0.316	0.041	-0.060	-0.041	-0.026

Esempio (continua)

Table 1b. Varimax rotated components

	Variance Explained by Rotated Components											
	1	2	3	4	5	6	7	8	9	10	11	12
	2.470	2.044	1.043	1.014	1.022	1.015	1.033	0.968	0.742	0.369	0.145	0.135
	Percent of Total Variance Explained											
	1	2	3	4	5	6	7	8	9	10	11	12
	20.585	17.037	8.690	8.452	8.517	8.457	8.610	8.070	6.181	3.073	1.206	1.121
	Rotated Loadings											
	1	2	3	4	5	6	7	8	9	10	11	12
Q	-0.293	0.868	-0.093	0.008	0.046	0.174	0.060	0.002	0.168	-0.029	-0.003	0.295
T	0.942	-0.070	0.010	0.088	0.090	-0.047	0.015	-0.173	0.008	0.068	0.235	0.021
pH	0.059	-0.021	-0.020	0.038	0.994	0.006	-0.065	-0.000	0.044	0.012	0.004	0.004
EC	0.725	-0.316	0.053	0.064	0.046	-0.024	0.032	-0.104	-0.072	0.589	0.018	0.001
SS	0.019	0.924	-0.018	0.100	-0.078	0.209	0.032	0.050	0.163	-0.085	-0.000	0.217
MA1	0.032	-0.285	0.054	0.013	-0.007	-0.950	-0.017	-0.011	-0.111	0.008	0.002	0.004
Cl	0.129	-0.062	0.958	-0.046	-0.021	-0.054	0.215	0.044	0.089	0.014	-0.008	0.005
NH ₃ -N	0.068	0.076	0.229	0.133	-0.080	0.017	0.938	0.086	0.159	0.009	-0.004	0.002
NO ₃ -N	-0.140	-0.079	0.043	-0.971	-0.041	0.011	-0.119	-0.114	-0.009	-0.017	0.002	0.003
DO	-0.906	0.028	-0.175	-0.097	0.022	-0.009	-0.075	0.192	-0.006	0.077	0.299	.008
Pv	-0.016	0.482	0.151	0.014	0.083	0.182	0.262	0.020	0.796	-0.033	-0.000	0.006
BOD ₅	-0.323	0.039	0.049	0.130	-0.000	0.012	0.088	0.930	0.015	-0.031	0.004	0.001

Ripensamento della ACP

La ACP può essere considerato una procedura ad effetti fissi in cui cioè la matrice dei dati osservata costituisce la "popolazione" di interesse.

Le conclusioni raggiunte sono valide sono limitate a tale specifico ambito.

La ACP descrive e spiega in che modo la variabilità totale può essere ripartita e quali sono i contenuti più rilevanti che si sono evidenziati.

Per meglio comprendere l'azione della ACP è bene rileggerla nel quadro della scomposizione spettrale delle matrici.

$$A \text{ is } m \times n$$

$$A = (\text{orthogonal}) (\text{diagonal}) (\text{orthogonal})$$

$$A = U \Sigma V^T$$

$$\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} = \begin{bmatrix} | & & & & | \\ U & & & & \\ | & & & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \ddots & \\ & & & & \sigma_r & & & & \end{bmatrix} \begin{bmatrix} | & & & & | \\ V^T & & & & \\ | & & & & | \end{bmatrix}$$

Equimax

Tenuto conto che i due precedenti criteri puntano a due diverse angolature di semplificazione è d'obbligo considerare anche metodi che tengano conto di entrambe.

Fra i tanti si può considerare l'Equimax che pone sullo stesso piano il criterio usato per il Quartimax e quelle del Varimax.

Equimax massimizza una somma ponderata dei criteri usati in varimax e quartimax cercando una struttura semplice sia per le righe che per le colonne.

Diagonalizzazione di una matrice

Data la matrice quadrata A di ordine mxm è possibile impostare la relazione

$$AU = UD$$

dove D è una matrice diagonale i cui elementi sono gli autovalori di A ed U una matrice formata con gli autovettori ad essi associati.

Se A è simmetrica di rango pieno, allora

$$A = QDQ^t, \quad D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m), \quad Q^t Q = Q Q^t = I$$

Con Q ortonormale.

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}; \lambda_1 = 5, \lambda_2 = -1, \lambda_3 = -1$$

E' facile verificare che Q è ortonormale e che QAQ=D=diag(5,-1,-1).

$$U = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \Rightarrow Q = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -2 & 0 \\ \sqrt{2} & 1 & -\sqrt{3} \\ \sqrt{2} & 1 & \sqrt{3} \end{bmatrix}$$

Scomposizione spettrale

Una matrice A simmetrica di ordine $m \times m$ può essere scritta come somma di prodotti esterni di vettori ortonormali e di norma unitaria

$$A = \sum_{j=1}^m \lambda_j \mathbf{u}_j \mathbf{u}_j^t$$

Dove λ_j è un autovalore di A e \mathbf{u}_j è l'autovettore normalizzato associato

Applichiamo questa proprietà alla matrice identità $\mathbf{I} = \sum_{j=1}^m \mathbf{u}_j \mathbf{u}_j^t$

Premoltiplichiamo entrambi i lati per A per dimostrare che

$$A\mathbf{I} = A \left(\sum_{j=1}^m \mathbf{u}_j \mathbf{u}_j^t \right) \Rightarrow A = \sum_{j=1}^m A \mathbf{u}_j \mathbf{u}_j^t \Rightarrow A = \sum_{j=1}^m \lambda_j \mathbf{u}_j \mathbf{u}_j^t \mathbf{u}_j \mathbf{u}_j^t = \sum_{j=1}^m \lambda_j \mathbf{u}_j \mathbf{u}_j^t$$

Altro esempio

$$A = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix} \quad \lambda_1 = \lambda_2 = 9, \lambda_3 = 18$$

$$\mathbf{e}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ -\frac{4}{\sqrt{18}} \end{pmatrix} \quad \mathbf{e}_3 = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$$

Note that since $\lambda_1 = \lambda_2$ the labeling of \mathbf{e}_1 and \mathbf{e}_2 is arbitrary.

- The lengths: $\mathbf{e}_1^t \mathbf{e}_1 = \mathbf{e}_2^t \mathbf{e}_2 = \mathbf{e}_3^t \mathbf{e}_3 = 1$.
- Orthogonality: $\mathbf{e}_1^t \mathbf{e}_2 = \mathbf{e}_1^t \mathbf{e}_3 = \mathbf{e}_2^t \mathbf{e}_3 = 0$.
- Decomposition:

$$A = 9\mathbf{e}_1 \mathbf{e}_1^t + 9\mathbf{e}_2 \mathbf{e}_2^t + 18\mathbf{e}_3 \mathbf{e}_3^t$$

Esempio

Matrix A is decomposed into p ($p \times p$) component matrices. where $\mathbf{e}_i^t \mathbf{e}_i = 1$ for all i , and $\mathbf{e}_i^t \mathbf{e}_j = 0$ for all $i \neq j$.

$$A = \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} \quad \lambda_1 = 6 \quad \mathbf{e}_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$$\lambda_1 \mathbf{e}_1 \mathbf{e}_1^t + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^t = 6 \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} - 4 \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} = A$$

Da notare che la scomposizione spettrale o scomposizione in valori singolari può essere realizzata anche per matrici rettangolari, pur con distinte matrici che premoltiplicano e postmoltiplicano la matrice data.

Altro esempio (continua)

$$A = 9 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} + 9 \begin{pmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ -\frac{4}{\sqrt{18}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & -\frac{4}{\sqrt{18}} \end{pmatrix}$$

$$+ 18 \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{9}{2} & \frac{9}{2} & 0 \\ \frac{9}{2} & \frac{9}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{9}{18} & -\frac{9}{18} & -\frac{36}{18} \\ -\frac{9}{18} & \frac{9}{18} & \frac{36}{18} \\ -\frac{36}{18} & \frac{36}{18} & \frac{144}{18} \end{pmatrix} + \begin{pmatrix} \frac{72}{9} & -\frac{72}{9} & \frac{36}{9} \\ -\frac{72}{9} & \frac{72}{9} & -\frac{36}{9} \\ \frac{54}{9} & \frac{36}{9} & \frac{18}{9} \end{pmatrix}$$

$$= \frac{1}{18} \begin{pmatrix} 234 & -72 & 36 \\ -72 & 234 & -36 \\ 36 & -36 & 180 \end{pmatrix} = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix}$$

Norma di Frobenius

Partiamo da una delle tante funzioni che misurano la norma di una matrice, la norma di Frobenius

$$\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2} = \text{Traccia}(A^t A)$$

$$4 + 4 + 16 + 4 + 1 + 9 + 0 + 1 + 2.25 + 2.25 + 16 + 9 + 2.25 + 2.25 = 48$$

$$X^t X = \begin{bmatrix} 2 & 2 & 0 & -1 \\ 2 & 1 & -1.5 & 1.5 \\ -4 & -3 & 1.5 & 1.5 \end{bmatrix} \begin{bmatrix} 2 & 2 & -4 \\ 2 & 1 & -3 \\ 0 & -1.5 & 1.5 \\ -1 & -0.5 & 1.5 \end{bmatrix} = \begin{bmatrix} 9 & 4.5 & -15.5 \\ 4.5 & 9.5 & -11 \\ -15.5 & -11 & 29.5 \end{bmatrix} \Rightarrow \text{Traccia} = 9 + 9.5 + 29.5 = 48$$

Poiché la nostra matrice dei dati è centrata la traccia di $(X^t X)$ corrisponde alla somma delle devianze delle variabili nella colonna dei dati

Soluzione di Eckart e Young

Nel caso di pesi uguali la soluzione si ottiene con la scomposizione in valori singolari della matrice dei dati centrata

$$\hat{X} = \sum_{r=1}^p \lambda_r (z_r v_r^t), \quad v_r^t v_s = z_r^t z_s = \begin{cases} 1 & \text{se } r = s \\ 0 & \text{se } r \neq s \end{cases}$$

Con $\lambda_r, r=1,2,\dots,p$ sono gli autovalori di \hat{X} disposti in ordine decrescente di grandezza e dove z_r e v_r sono gli autovettori di riga e di colonna ortogonali

Ovvero

$$\hat{X} = \underset{n \times p}{Z} \underset{p \times p}{D} \underset{p \times m}{V^t} \quad \text{dove } D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} \quad \text{e con } V^t V = Z^t Z = I_p$$

Problema di Eckart e Young

E' il problema della approssimazione della matrice dei dati con una matrice che ha un rango inferiore a quello della matrice originale.

Si tratta cioè di trovare un'altra matrice Y di rango prefissato, diciamo p, che renda minima la perdita di informazioni, misurata con la norma di Frobenius.

$$\underset{Y}{\text{Minimo}} \{ \|X - Y\| \}, \quad \text{soggetto a } \text{rango}(Y) = p$$

In una forma più generale si può porre

$$\underset{Y: \text{rango}(Y)=p}{\text{Minimo}} \left\{ \sqrt{\sum_{i=1}^n \sum_{j=1}^m w_{ij} (\hat{x}_{ij} - g_i^t h_j)^2} \right\}$$

Dove gli w_{ij} sono dei pesi che quantificano il grado di importanza da dare a ciascuno degli scarti

ACP e SVD

La scomposizione in valori singolari della matrice di varianze e covarianze ovvero della matrice di correlazione corrisponde al calcolo delle componenti principali delle rispettive matrici

$$W = \sum_{r=1}^p \lambda_r (u_r u_r^t), \quad u_r^t u_s = \begin{cases} 1 & \text{se } r = s \\ 0 & \text{se } r \neq s \end{cases}$$

Con $\lambda_r, r=1,2,\dots,p$ sono gli autovalori di W

Forse non è inopportuno rilevare che la SVD è anche un metodo di calcolo degli autovalori e degli autovettori.



ACP

Represent a grayscale $m \times n$ image by an $m \times n$ matrix A . (Requires mn numbers to represent.)

Find a low-rank matrix B that is as close as possible to A . (For rank r , requires only $r(m + n)$ numbers to represent.)



¿Donde estas la inferencia?/2

La ACP è nota da quasi 90 anni e si è molto cercato di dare un senso inferenziale ai ragionamenti condotti sulla matrice dei dati.

Secondo Jolliffe (Principal component analysis, 2nd ed.) 2002 c'è ormai poco da aggiungere a quello che è già noto:



La trattazione è complessa, a meno che non ci si limiti alle relazioni lineari.



Le poche certezze ottenute valgono solo per campioni estremamente grandi e sono inadeguate se non fuorvianti per campioni di dimensione ragionevole



Hanno come presupposto l'ipotesi che le variabili della popolazione siano multigaussiana che, come è noto, esistono forse solo nei libri di testo e nelle simulazioni.

Nei dati reali sono più rare del liocorno.

¿Donde estas la inferencia?

Supponiamo che X sia un vettore di variabili casuali gaussiane con media nulla e matrice di varianze covarianze W

$$X \sim N_m(0, W)$$

Sia data la scomposizione in valori singolari di W

$$W = QDQ^t$$

Le trasformazioni lineari $F=Q^tX$ sono distribuite secondo la gaussiana

$$F \sim N_m(0, D)$$

Questo perché le componenti sono state costruite in modo da risultare non correlate.

Inoltre la trasformazione lineare $F^*=D^{-0.5}Q^tX$ ha distribuzione

$$F^* \sim N_m(0, I)$$

Biplot

Consideriamo una matrice di dati con n righe ed m colonne avente rango almeno pari a due.

Il Bi-plot E' una tecnica grafica (plot) per rappresentare insieme nel piano sia le variabili che le unità di una matrice dei dati.

E' a questa simultaneità di rappresentazione che fa riferimento il prefisso "bi" nella denominazione di Biplot.

$$X = \begin{array}{c|cccc} & V_1 & V_2 & V_j & V_m \\ \hline U_1 & x_{11} & x_{12} & x_{1j} & x_{1m} \\ U_2 & x_{21} & x_{22} & x_{2j} & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_i & x_{i1} & x_{i2} & x_{ij} & x_{im} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_n & x_{n1} & x_{n2} & x_{in} & x_{nm} \end{array}$$

La stessa tecnica può essere estesa alla rappresentazione tridimensionale rimanendo "bi".



Prof. Ruben Gabriel, "The founder of biplot"

Any two-way table can be analyzed using a 2D-biplot as soon as it can be sufficiently approximated by a rank-2 matrix." (Gabriel, 1971)

Biplot/2

Convienne centrare la matrice dei dati

$$\hat{X} = CX \quad \text{con} \quad C = \left(I_n - \frac{1}{n} uu^t \right)$$

Con il centramento si eliminano le differenze relative al diverso livello delle variabili che ora sono forzate ad avere tutte lo stesso livello pari a zero.

Rimangono le differenze dovute alle scale di misurazione.

A questo fine, i dati sono sottoposti ad altre trasformazioni quali la standardizzazione, la unitarizzazione oppure la trasformazione logaritmica

La tecnica del Biplot è inappropriata se nella matrice dei dati sono presenti variabili binarie, politome sconnesse e gli ordinamenti a meno che questi ultimi non siano ad almeno 7 livelli.

$$\begin{bmatrix} 2 & 2 & -4 \\ 2 & 1 & -3 \\ 0 & -1.5 & 1.5 \\ -1 & -0.5 & 1.5 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 1 \\ 0 & -1.5 \\ -1 & -0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

Esempio

Scegliamo $p=2$.

$$= \begin{bmatrix} (2 \ 2) \\ (2 \ 1) \\ (0 \ -1.5) \\ (-1 \ -0.5) \end{bmatrix} \begin{bmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 2*1+2*0 & 2*0+2*1 & 2*(-1)+2*(-1) \\ 1*2+1*0 & 2*0+1*1 & 2*(-1)+1*(-1) \\ 0*0+(-1.5)*0 & 0*0+(-1.5)*1 & 0*(-1)+(-1.5)*(-1) \\ (-1)*1+(-0.5)*0 & (-1)*0+(-0.5)*1 & (-1)*(-1)+(-0.5)*(-1) \end{bmatrix} = \begin{bmatrix} 2 & 2 & -4 \\ 2 & 1 & -3 \\ 0 & -1.5 & 1.5 \\ -1 & -0.5 & 1.5 \end{bmatrix}$$

La fattorizzazione appena proposta però non è unica. Infatti:

$$\begin{bmatrix} 2 & 2 & -4 \\ 2 & 1 & -3 \\ 0 & -1.5 & 1.5 \\ -1 & -0.5 & 1.5 \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ 2 & -1 \\ -3 & 4.5 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} -3 & -1 & 4 \\ -2 & -1 & -3 \end{bmatrix}$$

Fattorizzazione della matrice dei dati

Ogni matrice di rango maggiore o uguale a 2 può essere espressa come il prodotto di due altre matrici

$$\hat{X}_{n \times m} = G_{n \times p} H_{p \times m}$$

Che possiamo esprimere con maggiore dettaglio

$$\hat{X}_{n \times m} = [g_1 \ g_2 \ \dots \ g_n] \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix}$$

La matrice G è definita da n vettori (1 x p) riferibili ai marker (cioè punti).

La matrice H è definita da m vettori (p x 1) riferibili ai marker delle variabili.

Quindi la matrice dei dati può essere sintetizzata da (n + m) marker.

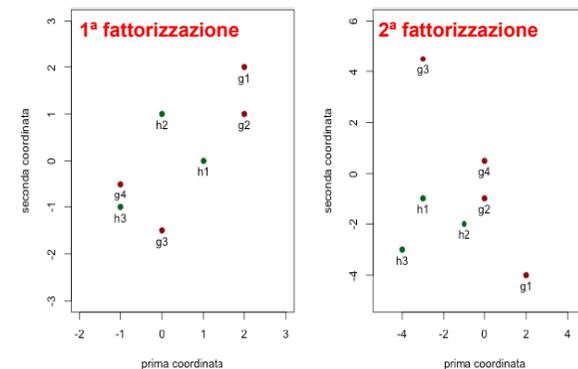
Ad esempio, se la matrice dei dati è relativa a 100 unità per 20 variabili, i marker sarebbero 120

Scatterplot

I grafici evidenziano il collocamento relativo dei vettori unità e dei vettori variabile.

Con $p=2$ possiamo visualizzare i marker delle unità e delle variabili in un piano.

Possiamo anche operare in 3D.



C'è comunque una notevole differenza tra le due rappresentazioni; tuttavia la matrice prodotto che si ricava è la stessa.

In effetti, sono infinite le coppie di matrici che possono portare ad una matrice prefissata.

Non identificabilità

Supponiamo che Q sia una qualsiasi matrice non singolare cioè tale che Q^{-1} esista.

Se post-moltiplichiamo la matrice nel primo fattore G per Q e, nello stesso tempo, pre-moltiplichiamo la matrice nel secondo fattore H per l'inversa Q^{-1} si ottiene

$$X = (GQ)(Q^{-1}H) = GQQ^{-1}H = GIH = GH$$

Alcune scelte delle matrici G ed H privilegiano la messa in evidenza delle unità, altre scelte quelle delle variabili.

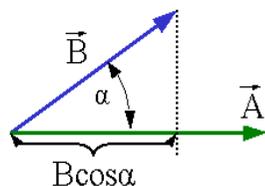
Sarà necessario trovare una opzione di equilibrio.

Geometria del prodotto scalare

L'espressione di x_{ij} come prodotto scalare facilita la sua interpretazione geometrica.

Infatti, per la nota legge dei coseni, si ha

$$g_i^t h_j = \|g_i\| \|h_j\| \cos(\alpha_{ij})$$



$$\vec{A} \cdot \vec{B} = |\vec{A}| |\vec{B}| \cos \alpha$$

Dove α_{ij} è l'angolo formato dai vettori g_i e h_j .

Ne consegue che, la lettura del grafico Biplot deve tenere conto sia della lunghezza dei vettori marker che degli angoli che essi formano.

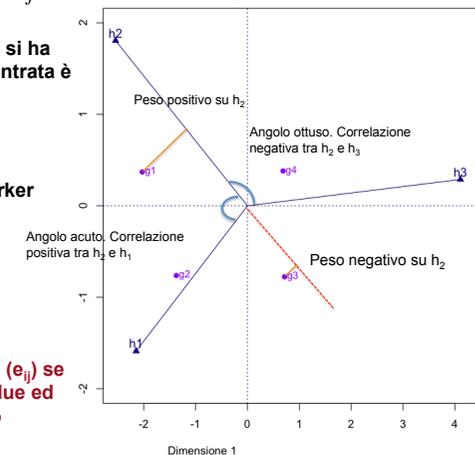
Relazione tra i marker

La relazione tra i vettori delle matrici a fattore e gli elementi originali della matrice dei dati (eventualmente trasformata) è data dalla equazione

$$\hat{x}_{ij} = g_i^t h_j + e_{ij} \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m$$

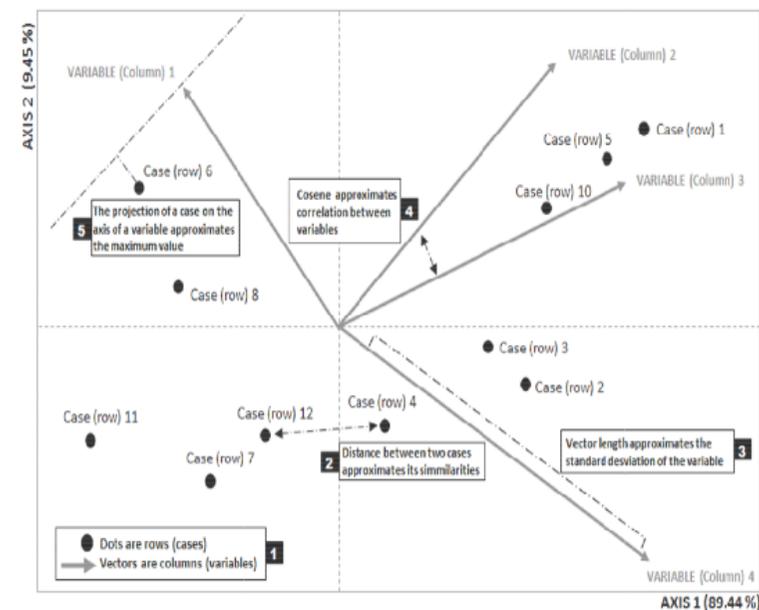
dove e_{ij} esprime il deficit che si ha se il rango della matrice X centrata è maggiore di due.

Quindi ogni \hat{x}_{ij} equivale al prodotto scalare del marker dell'unità i -esima g_i per il marker della j -esima variabile h_j .



La rappresentazione è esatta (e_{ij}) se la matrice dei dati ha rango due ed è approssimata se è di rango superiore a due.

Biplot in pratica



Biplot e SVD

La tecnica del Biplot ritiene di poter rappresentare il contenuto informativo della matrice dei dati in $p=2$ o $p=3$ dimensioni.

A questo si usa la scomposizione in valori singolari che approssimazioni di rango via via crescente ad una matrice data in base alla norma di Frobenius.

$$Y_2 = \lambda_1 z_1 v_1^t + \lambda_2 z_2 v_2^t; \quad Y_3 = \lambda_1 z_1 v_1^t + \lambda_2 z_2 v_2^t + \lambda_3 z_3 v_3^t$$

$$Y = [z_1, \dots, z_q] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_q \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_q \end{bmatrix}, \quad W^t W = Z^t Z = I_q; \quad q = 2 \text{ o } 3$$

E' utile osservare che la scomposizione in valori singolari può anche esprimersi con la relazione moltiplicativa seguente

$$\hat{X} = ZD^\alpha D^{1-\alpha} V^t = (ZD^\alpha)(VD^{1-\alpha})^t \quad \text{dove} \quad D^\alpha = \begin{bmatrix} \lambda_1^\alpha & & \\ & \ddots & \\ & & \lambda_p^\alpha \end{bmatrix}$$

Calcolo delle coordinate/2

Se si pone $\alpha = 0$ allora si realizza il biplot-Gabriel

$$\begin{cases} g_i^t = (z_{1i}, \dots, z_{qi}), & i = 1, 2, \dots, n \\ h_j^t = (\lambda_1 v_{1j}, \dots, \lambda_q v_{qj}), & j = 1, 2, \dots, m \end{cases} \Rightarrow G = Z \quad e \quad H = DV^t$$

Ne consegue che i marker delle unità hanno tutti norma unitaria, mentre i marker delle variabili avrebbero lunghezza diversa. La lunghezza degli assi approssima la varianza.

Se si pone $\alpha = 1$ allora si realizza il biplot JK (jeikei)

$$\begin{cases} g_i^t = (\lambda_1 z_{1i}, \dots, \lambda_q z_{qi}), & i = 1, 2, \dots, n \\ h_j^t = (v_{1j}, \dots, v_{qj}), & j = 1, 2, \dots, m \end{cases} \Rightarrow G = ZD \quad e \quad H = V^t$$

Qui la norma dei marker delle variabili è sempre pari ad uno laddove i marker delle unità avrebbero lunghezza variabile.

Calcolo delle coordinate

La presentazione col Biplot di una matrice dei dati si semplifica se si parte dalla matrice di varianze-covarianze.

Se però le variabili hanno scale di misurazioni eterogenee o hanno campi di variazioni molto diversi sarà opportuno realizzare il Biplot in base alla matrice di correlazione R.

La determinazione delle matrici G e H si baserà sulla scomposizione

$$R_p = \sum_{r=1}^p z_r (\lambda_r^\alpha \lambda_r^{1-\alpha}) v_r^t = \sum_{r=1}^p \left(\lambda_r^\alpha z_r \right) \left(\lambda_r^{1-\alpha} v_r \right)^t$$

Effetto riga Effetto colonna

Le matrici delle coordinate sono ora

p è il rango della \hat{X} e q il numero di dimensioni (2 o 3)

$$\begin{cases} g_i^t = (\lambda_1^\alpha z_{1i}, \dots, \lambda_q^\alpha z_{qi}), & i = 1, 2, \dots, n \\ h_j^t = (\lambda_1^{1-\alpha} z_{1j}, \dots, \lambda_q^{1-\alpha} z_{qj}), & j = 1, 2, \dots, m \end{cases} \Rightarrow G = ZD^\alpha, \quad H = VD^{1-\alpha}$$

La scelta dell'esponente α caratterizzerà il grafico biplot

Calcolo delle coordinate/3

Una soluzione intermedia che garantisce alle coordinate delle unità e delle variabili la stessa lunghezza e quindi pesi comparabili per entrambi gli aspetti del Biplot è $\alpha = 0.5$ (biplot-radice (sqrt))

$$\begin{cases} g_i^t = (\sqrt{\lambda_1} z_{1i}, \dots, \sqrt{\lambda_q} z_{qi}), & i = 1, 2, \dots, n \\ h_j^t = (\sqrt{\lambda_1} v_{1j}, \dots, \sqrt{\lambda_q} v_{qj}), & j = 1, 2, \dots, m \end{cases} \Rightarrow G = ZD^{0.5}, \quad H = VD^{0.5}$$

con la conseguenza che

$$G^t G = (ZD^{0.5})^t (ZD^{0.5}) = D^{0.5} Z^t Z D^{0.5} = D^{0.5} I D^{0.5} = D$$

$$H^t H = (VD^{0.5})^t (VD^{0.5}) = D^{0.5} V^t V D^{0.5} = D^{0.5} I D^{0.5} = D$$

In entrambi i casi la matrice prodotto non è l'identità e quindi si viola il requisito di univocità

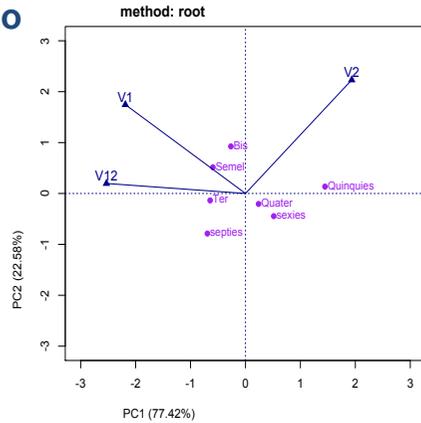
Ordine	V1	V2	V12
Semel	2	0	2
Bis	2	1	1
Ter	1	-1	2
Quater	-1	0	-1
Quinquies	-3	2	-5
sexies	-2	0	-2
septies	0	-2	2

Esempio

```
> svd(Xh)
$d
[1] 7.753161 3.510727 1.887670

$u
      [,1] [,2] [,3]
[1,] 0.08775642 0.675766394 0.7318733
[2,] -0.19303955 0.499569853 -0.4381251
[3,] -0.31299502 -0.075455575 0.1072012
[4,] 0.16867819 -0.141371426 0.1103080
[5,] 0.79466824 0.009539724 -0.1040943
[6,] -0.33735638 0.282742852 -0.2206159
[7,] -0.28863367 -0.433654002 0.4350182

$V
      [,1] [,2] [,3]
[1,] -0.4837971 0.8412660 0.2412714
[2,] 0.2949196 0.4162708 -0.8600820
[3,] -0.8239919 -0.3449496 -0.4494965
```



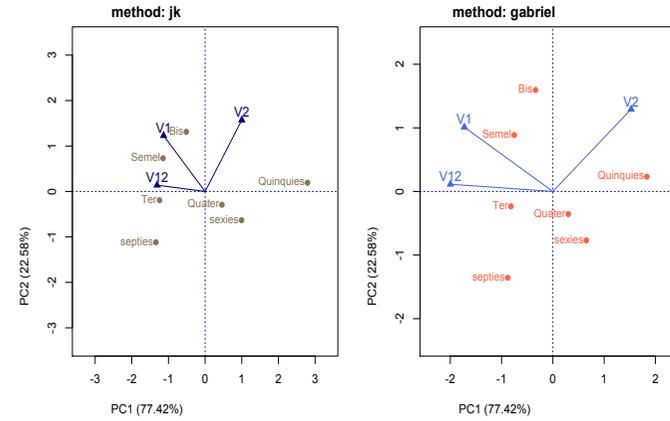
Ogni combinazione di vettori presenti nel Biplot rappresenta un dato della matrice fattorizzata.

I punti ravvicinati sono punti che presentano punteggi (scores) simili sulle componenti o assi.

Maggiore è la variabilità spiegata maggiore è la corrispondenza tra la distanza nel biplot e la distanza nella matrice originale dei dati.

Biplot JK e Biplot root

Con piccoli data set è difficile che si modifichino le conclusioni al modificarsi del metodo biplot.



Le variabili puntano verso la componente (asse) con il quale sono più correlate.

La norma del vettore è proporzionale al quadrato della correlazione tra i valori della variabile ed i valori interpolati a partire dalle componenti sugli assi.

Qualità della rappresentazione

Il grado di approssimazione ottenuto dal biplot è facilmente quantificabile. Ad esempio con:

$$\beta_q = \text{Gabriel} = \frac{\frac{2}{q} \sum_{r=1}^q \lambda_r^2}{\sum_{r=1}^m \lambda_r^2}$$

$$JK = \frac{\sum_{r=1}^q \lambda_r^2}{\sum_{r=1}^m \lambda_r^2} = \frac{2}{q}$$

NB. Il Biplot si può applicare anche se le variabili sono più numerose delle unità

Se è vicina ad uno, l'approssimazione è soddisfacente (se il rango della matrice dei dati è p=q, allora l'approssimazione è perfetta).

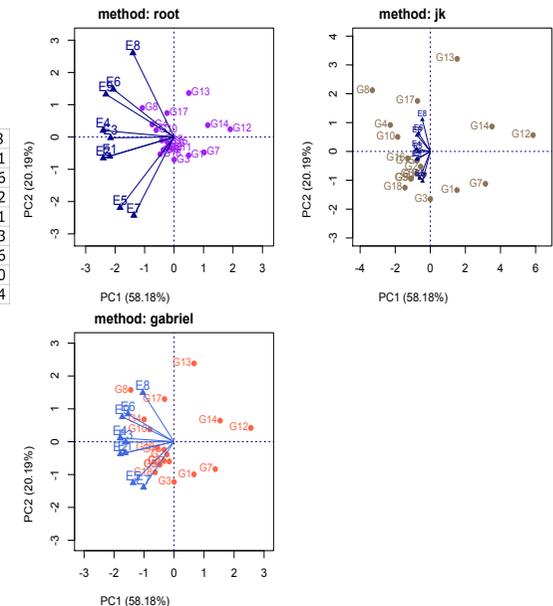
Se si avvicina allo zero (minore di p/m) la riproduzione è inadeguata ed il biplot non potrà essere di grande aiuto.

Altro esempio

Dati ambientali

Campioni	E1	E2	E3	E4	E5	E6	E7	E8
Q1	1	2	2	2	1	1	0	1
Q2	2	2	3	4	2	3	4	6
Q3	2	2	1	2	3	2	1	2
Q4	2	4	3	1	1	4	2	1
Q5	3	2	2	3	4	3	3	3
Q6	0	0	0	0	3	4	5	6
Q7	6	5	4	3	0	0	0	0
Q8	3	3	4	4	3	3	4	4

	E1	E2	E3	E4	E5	E6	E7	E8
E1	1.00	0.83	0.74	0.57	-0.38	-0.60	-0.43	-0.51
E2	0.83	1.00	0.84	0.37	-0.66	-0.47	-0.57	-0.72
E3	0.74	0.84	1.00	0.71	-0.48	-0.32	-0.22	-0.34
E4	0.57	0.37	0.71	1.00	0.05	-0.32	-0.01	0.06
E5	-0.38	-0.66	-0.48	0.05	1.00	0.56	0.64	0.60
E6	-0.60	-0.47	-0.32	-0.32	0.56	1.00	0.81	0.64
E7	-0.43	-0.57	-0.22	-0.01	0.64	0.81	1.00	0.92
E8	-0.51	-0.72	-0.34	0.06	0.60	0.64	0.92	1.00



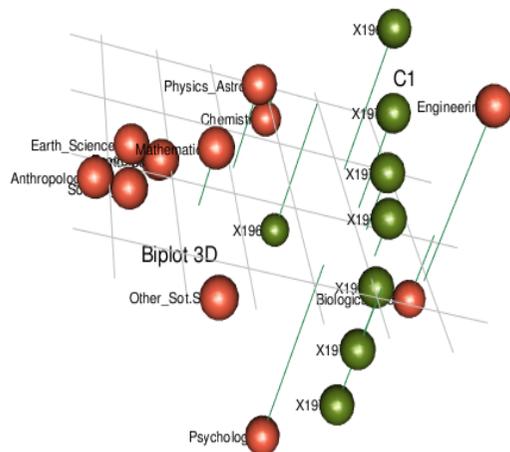
Biplot in 3D

Dottorati di ricerca conseguiti negli USA in varie discipline negli anni 1960, 1965, 1970, 1971-1975

Quality_3: 99.9%

In questo caso due dimensioni erano sufficienti

Quality_2: 98.3%



Altro esempio di Biplot in 3D

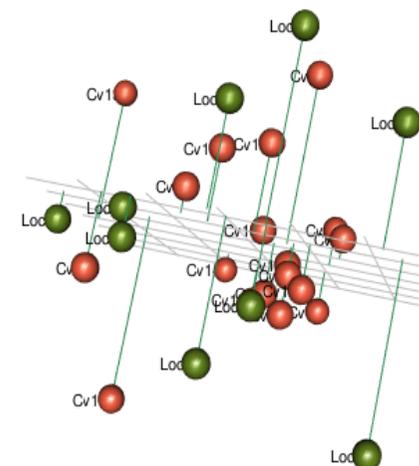
The sample data are yields from the 1993 Ontario winter wheat (*Triticum aestivum* L.) performance trials, in which 18 cultivars were tested at nine locations (Yan and Kang 2003)

Quality_3: 85.8%

Quality_2: 78.4%

La lettura del grafico si concentra nella valutazione delle distanze unità-unità e variabile-variabile.

In un modo meno semplice si può provare a valutare le relazioni unità-variabili.



Biplot HJ

Si parte dalla SVD della matrice dei dati

$$\hat{X} = \underset{n \times p}{Z} \underset{p \times p}{D} \underset{p \times m}{V}^t$$

E' come il metodo "root" solo che l'esponente comune non è 0.5, ma 1.

Galindo (1985) ha proposto la fattorizzazione

$$\begin{cases} g_i^t = (\lambda_1 z_{1i}, \dots, \lambda_q z_{qi}), & i = 1, 2, \dots, n \\ h_j^t = (\lambda_1 v_{1j}, \dots, \lambda_q v_{qj}), & j = 1, 2, \dots, m \end{cases} \Rightarrow G = ZD, \quad H = VD$$

Ecco alcune relazioni che legano le due matrici della SVD

- 1) $\hat{X}V = ZDV^tV = ZD$
- 2) $Z^t\hat{X} = Z^tZDV^t = DV^t \Rightarrow \hat{X}^tZ = DV^t$
- 3) $G^tG = H^tH = D^2$

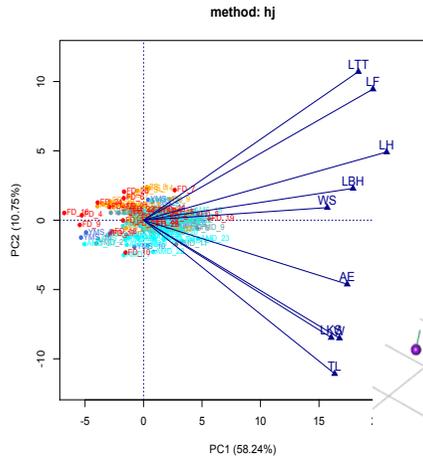
Esempio: sparrows data set

H. Bumpus (1899). Caratteristiche fisiche di passeri sopravvissuti ad un inverno molto severo.

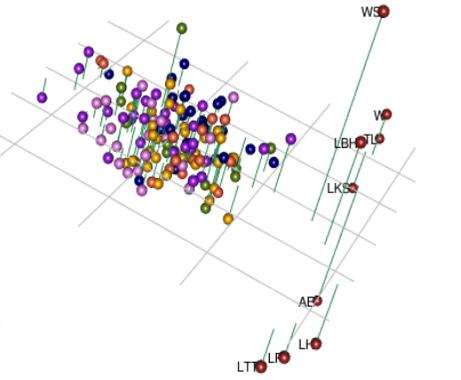


Variable	Male values	Female values	t-value ²
Total length	158.6 ± 3.6	156.3 ± 3.4	2.19*
Skull length	31.87 ± 0.57	31.69 ± 0.62	1.01
Skull width	15.65 ± 0.28	15.34 ± 0.43	2.78**
Tarsometatarsus length	19.63 ± 0.65	19.60 ± 0.89	-1.37
Tibiotarsus length	28.97 ± 0.76	28.37 ± 1.03	2.17*
Femur length	17.71 ± 0.47	17.37 ± 0.65	1.96
Wing length	77.21 ± 1.59	73.79 ± 1.75	6.71***
Humerus length	18.91 ± 0.49	18.39 ± 0.59	3.16**
Keel length	21.87 ± 0.75	20.63 ± 1.03	4.31***
Fresh body mass (g)	29.41 ± 0.86	28.49 ± 1.72	2.19*
Lipid content as %			
fresh body mass	7.41 ± 0.66	8.17 ± 1.33	2.35*
Crop and proventriculus content (g dry mass)	0.26 ± 0.14	0.26 ± 0.23	-0.04

Sparrows/2



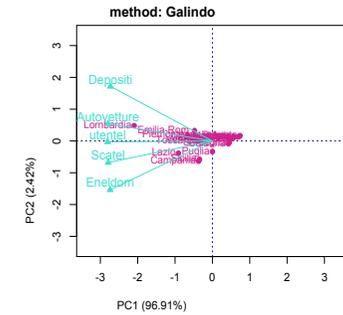
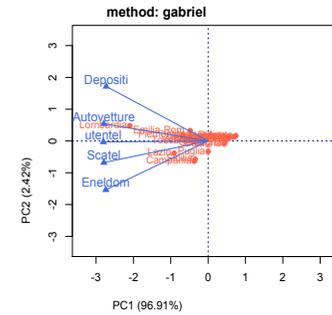
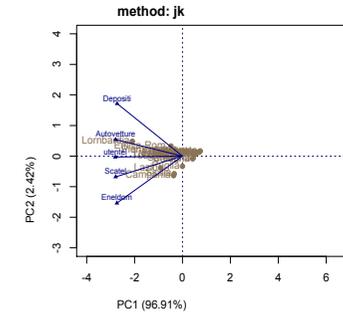
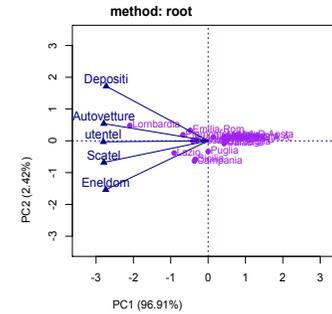
The Bumpus data generally agree that females suffered proportionately greater mortality than males and that female survivorship reflects stabilizing or normalizing selection.



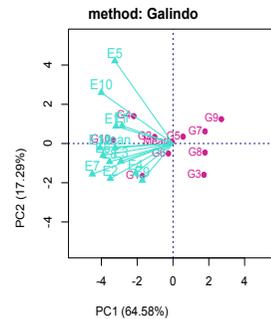
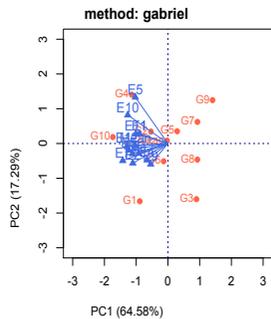
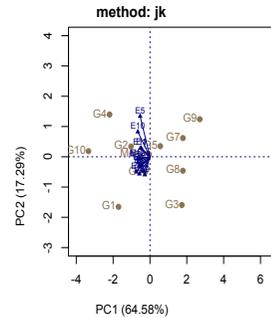
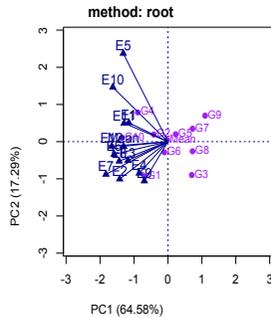
Quality_3: 76.6%

Disagreement persists, however, in determining whether male survivorship reflects directional selection and, if so, whether this selects for larger versus smaller individuals

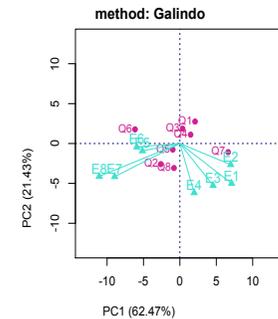
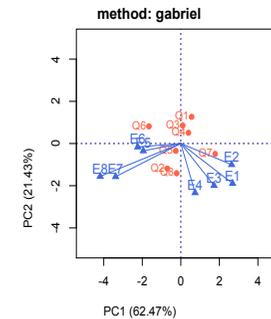
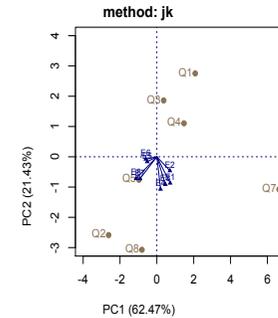
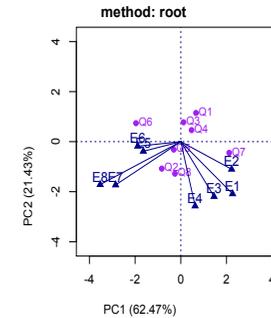
Regioni: indicatori benessere



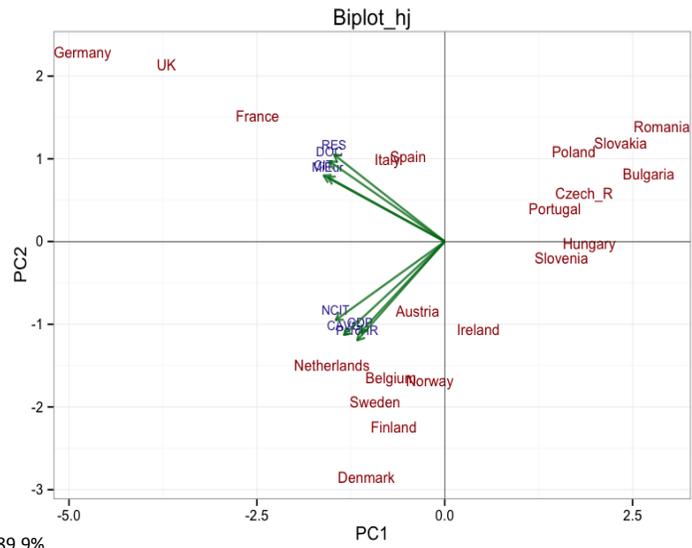
Test environments



Artif: dati non standardizzati



Scientific effort and bibliometrics indicators for European Countries



Top 25 universities according to 4 Ranking variables

