

# Appunti di Analisi Fattoriale

Analisi multivariata, Corso di laurea in Statistica. Carla Rampichini

## 1 Rotazioni

Abbiamo detto che può essere utile ruotare il sistema di coordinate di un certo angolo  $\theta$  per cercare una migliore interpretazione dei fattori. In particolare, mostreremo che scegliendo una rotazione opportuna si può ottenere una matrice dei *factor loadings*  $\mathbf{L}\mathbf{T}$  più semplice da interpretare. I *loadings* rappresentano le covarianze (correlazioni) tra i fattori e le variabili originali (standardizzate), quindi ha senso cercare le rotazioni che portano a dei fattori che hanno la massima correlazione con vari gruppi di variabili originali. Come si può ruotare il sistema degli assi?

Quando consideriamo un punto  $\mathbf{x}$  in  $\mathbb{R}^p$ , per ottenere la sua rappresentazione geometrica utilizziamo un sistema di coordinate  $p$  dimensionale. Consideriamo per esempio il punto  $P$  di coordinate  $\mathbf{x} = (x_1, x_2)'$  in  $\mathbb{R}^2$  rispetto ad un dato sistema di assi ortogonali.

Sia  $\mathbf{T}$  una matrice ortogonale  $2 \times 2$

$$\mathbf{T}(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (1)$$

Se si ruotano gli assi di un angolo  $\theta$  in senso orario, le nuove coordinate del punto  $P$  sono date dal vettore  $\mathbf{y} = \mathbf{T}\mathbf{x}$ . mentre una rotazione dello stesso angolo in senso anti-orario fornisce le nuove coordinate  $\mathbf{y} = \mathbf{T}'\mathbf{x}$ .

Più in generale, pre-moltiplicare un vettore  $\mathbf{x}$  per una matrice ortogonale  $\mathbf{T}$  corrisponde geometricamente ad una rotazione degli assi, in modo tale che il primo nuovo asse sia determinato dalla prima riga di  $\mathbf{T}$ .

### 1.1 Rotazione dei fattori

Per stimare il modello fattoriale con il metodo della massima verosimiglianza (ML) abbiamo imposto il vincolo  $\mathbf{L}'\Psi^{-1}\mathbf{L} = \Delta$  diagonale. Questo vincolo è utile da un punto di vista analitico, per trovare una soluzione unica, ma può rendere più complicata l'interpretazione dei parametri. L'interpretazione dei *loadings* risulta molto più semplice se le variabili possono essere suddivise in sottoinsiemi ognuno associato ad un unico fattore.

Un algoritmo di rotazione dei fattori molto utilizzato è il c.d. metodo **varimax** proposto da Kaiser (1985). Vediamo come si applica questo algoritmo nel caso semplice di due soli fattori ( $m = 2$ ).

Sia  $\mathbf{T}(\theta)$  la matrice ortogonale (1). Utilizziamo  $\mathbf{T}$  per ruotare gli assi fattoriali in senso orario di un angolo pari a  $\theta$ . La rotazione corrispondente per i *loadings* è calcolata come  $\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\mathbf{T}(\theta)$ . L'idea del *metodo varimax* consiste nel trovare l'angolo  $\theta$  che massimizza la somma delle varianze dei *loadings* al quadrato  $\ell_{kj}$  rispetto a ogni colonna di  $\widehat{\mathbf{L}}$ .

Definiamo  $\tilde{\ell}_{kj} = \widehat{\ell}_{kj}^*/\widehat{h}_j^*$ , il criterio varimax sceglie l'angolo di rotazione  $\theta$  in modo tale da massimizzare la quantità

$$\mathcal{V} = \frac{1}{p} \sum_{j=1}^m \left\{ \sum_{k=1}^p (\tilde{\ell}_{kj})^4 - \left[ \frac{1}{p} \sum_{k=1}^p (\tilde{\ell}_{kj})^2 \right]^2 \right\}$$

## 1.2 Punteggi fattoriali

Fino ad ora abbiamo visto strategie di costruzione del modello basate sulla stima e interpretazione dei factor loadings e delle comunaltà. Questo è il primo passo nella costruzione del modello, in quanto i fattori  $\mathbf{F}$  sono considerati fattori comuni generici (normalizzati) che consentono di spiegare l'associazione tra le variabili originali.

I valori stimati dei fattori comuni sono detti *punteggi fattoriali* (factor scores), e sono utili sia per interpretare i risultati che per fare diagnostica. I punteggi fattoriali sono previsioni dei fattori latenti calcolate per ogni individuo  $i$ ,  $i = 1, \dots, n$ . Johnson and Wichern descrivono tre metodi three methods per calcolare i punteggi fattoriali che in pratica portano a risultati simili. Noi considereremo solo il metodo della regressione che ha il vantaggio di essere il più semplice ed è facile da implementare.

## Bibliografia

Johnson R.A. e Wichern D.W. (2007). *Applied Multivariate Statistical Analysis*. Sixth Edition. Pearson Education International.

Kaiser, H. F. (1985). The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, 23, 187-200.