CdL in Scienze Statistiche ed Economiche - Università degli Studi di Milano-Bicocca

Lezione : Analisi fattoriale

Docente: Aldo Solari

Nelle scienze sociali, in particolare in psicologia, spesso è problematico misurare le variabili di interesse direttamente. Ad esempio:

- Intelligenza
- Classe sociale

Queste variabili non osservabili (variabili latenti) sono chiamate *fattori comuni*. E' possibile esaminare queste variabili indirettamente, misurando variabili osservabili che sono ad esse collegate. Ad esempio

- Punteggio in varie prove di intelligenza, etc.
- Occupazione, Tasso di istruzione, Casa di proprietà, etc.

L'obiettivo dell'*analisi fattoriale* è studiare le relazioni tra variabili osservabili e fattori comuni

1 Il modello fattoriale con k fattori

$$\begin{array}{rcl} x_1 & = & \lambda_{11} f_1 + \ldots + \lambda_{1k} f_k + u_1 \\ x_2 & = & \lambda_{21} f_1 + \ldots + \lambda_{2k} f_k + u_2 \\ \vdots & = & \vdots \\ x_p & = & \lambda_{p1} f_1 + \ldots + \lambda_{pk} f_k + u_p \end{array}$$

dove

- $\underset{p\times 1}{x} = (x_1, \ldots, x_p)'$ sono le *variabili osservate* (variabili casuali)
- $\underset{k\times 1}{f} = (f_1, \ldots, f_k)'$ sono i *fattori comuni* (var. casuali non oss.)
- $\underset{p\times 1}{u} = (u_1, \ldots, u_p)'$ sono i *fattori specifici* (var. casuali non oss.)
- λ_{ij} sono i *pesi fattoriali* (costanti incognite)

In forma matriciale:

$$\underset{p\times 1}{x} = \underset{p\times k}{\Lambda}\underset{k\times 1}{f} + \underset{p\times 1}{u}$$

1.1 Assunzioni

1. Variabili osservate: $\mathbb{E}(\underset{p\times 1}{x}) = \underset{p\times 1}{0}$ (altrimenti centrare sullo 0)

2. Fattori comuni: $\mathbb{E}(\underset{k\times 1}{f}) = \underset{k\times 1}{0}$, $\mathbb{C}\text{ov}(\underset{k\times 1}{f}) = \mathbb{E}(\underset{k\times 1}{f}\underset{1\times k}{f'}) = \underset{k\times k}{I}$

3. Fattori specifici: $\mathbb{E}(\underset{p\times 1}{u}) = \underset{p\times 1}{0}$, $\mathbb{C}\text{ov}(\underset{p\times 1}{u}) = \mathbb{E}(\underset{p\times 1}{u}\underset{1\times p}{u'}) = \underset{p\times p}{\Psi} = diag(\psi_1, \ldots, \psi_p)$

4. Incorrelazione tra f e u: $\mathbb{C}\text{ov}(\underset{p\times 1}{u}, \underset{k\times 1}{f}) = \mathbb{E}(\underset{p\times 1}{u}\underset{1\times k}{f'}) = \underset{p\times k}{0}$

1.2 Proprietà

Proposition 1.1. *La matrice di varianza/covarianza Σ di x è data da*

$$\underset{p\times p}{\Sigma} = \underset{p\times k}{\Lambda}\underset{k\times p}{\Lambda'} + \underset{p\times p}{\Psi}$$

Dimostrazione.

$$\begin{aligned}
\underset{p\times p}{\Sigma} &= \mathbb{C}\text{ov}(\underset{p\times 1}{x}) = \mathbb{E}(\underset{p\times 1}{x}\underset{1\times p}{x'}) \\
&= \mathbb{E}[(\Lambda f + u)(\Lambda f + u)'] \\
&= \mathbb{E}[\Lambda f(\Lambda f)' + u(\Lambda f)' + (\Lambda f)u' + uu'] \\
&= \Lambda\mathbb{E}(ff')\Lambda' + \mathbb{E}(uf')\Lambda' + \Lambda\mathbb{E}(fu') + \mathbb{E}(uu') \\
&= \Lambda\mathbb{C}\text{ov}(f)\Lambda' + \mathbb{C}\text{ov}(u, f)\Lambda' + \Lambda\mathbb{C}\text{ov}(f, u) + \mathbb{C}\text{ov}(u) \\
&= \Lambda\Lambda' + \Psi
\end{aligned}$$

□

Il modello fattoriale con k fattori ipotizza che

$$p(p+1)/2$$

parametri corrispondenti alle p varianze e alle $p(p-1)/2$ covarianze di $\underset{p\times p}{\Sigma}$ possano essere espressi con

$$p(k+1)$$

parametri corrispondenti ai pk pesi fattoriali di $\underset{p\times k}{\Lambda}$ e le p varianze specifiche di $\underset{p\times p}{\Psi}$.
Per esempio, se abbiamo $p = 12$ variabili osservabili $\underset{p\times 1}{x}$ e un modello fattoriale con $k = 2$ fattori, allora i $p(p+1)/2 = 78$ parametri di $\underset{p\times p}{\Sigma}$ possono essere ridotti ai $p(k+1) = 36$ parametri di $\underset{p\times k}{\Lambda}$ e $\underset{p\times p}{\Psi}$.

La varianza di x_i si può esprimere come

$$\begin{aligned}\sigma_{ii} &= \mathbb{V}\mathrm{ar}(x_i) = \{\underset{p\times p}{\Sigma}\}_{ii} = \{\Lambda\Lambda'\}_{ii} + \{\Psi\}_{ii}\\ &= \sum_{j=1}^{k}\lambda_{ij}^2 + \psi_i\\ &= \underbrace{h_i^2}_{\text{comunalita}'} + \underbrace{\psi_i}_{\text{var. specifica}}\end{aligned}$$

dove

- $h_i^2 = \lambda_{i1}^2 + \ldots + \lambda_{ik}^2$ è la comunalità, ovvero la varianza dovuta ai k fattori comuni
- ψ_i è la varianza specifica di x_i non attribuibile ai fattori comuni

La covarianza tra x_i e x_j si può esprimere come

$$\begin{aligned}\sigma_{ij} &= \mathbb{C}\mathrm{ov}(x_i, x_j) = \{\underset{p\times p}{\Sigma}\}_{ij} = \{\Lambda\Lambda'\}_{ij} + \{\Psi\}_{ij}\\ &= \sum_{l=1}^{k}\lambda_{il}\lambda_{jl}\\ &= \lambda_{i1}\lambda_{j1} + \ldots + \lambda_{ik}\lambda_{jk}\end{aligned}$$

La covarianza tra x e f si può esprimere come

$$\begin{aligned}\mathbb{C}\mathrm{ov}(\underset{p\times 1}{x}, \underset{k\times 1}{f}) &= \mathbb{E}(\underset{p\times 1}{x}\underset{1\times k}{f'})\\ &= \mathbb{E}[(\Lambda f + u)f']\\ &= \Lambda\mathbb{E}(ff') + \mathbb{E}(uf')\\ &= \underset{p\times k}{\Lambda}\end{aligned}$$

quindi il peso fattoriale λ_{ij} rappresenta la covarianza tra x_i e f_j:

$$\mathbb{C}\mathrm{ov}(x_i, f_j) = \{\underset{p\times k}{\Lambda}\}_{ij} = \lambda_{ij}$$

Example 1.2. *Verifica della relazione* $\Sigma = \Lambda\Lambda' + \Psi$ *con* $k = 2$

Example 9.1 (Verifying the relation $\Sigma = LL' + \Psi$ for two factors) Consider the covariance matrix

$$\Sigma = \begin{bmatrix} 19 & 30 & 2 & 12\\ 30 & 57 & 5 & 23\\ 2 & 5 & 38 & 47\\ 12 & 23 & 47 & 68\end{bmatrix}$$

The equality

$$\begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

or

$$\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$$

may be verified by matrix algebra. Therefore, $\mathbf{\Sigma}$ has the structure produced by an $m = 2$ orthogonal factor model. Since

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \ell_{31} & \ell_{32} \\ \ell_{41} & \ell_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix},$$

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

the communality of X_1 is, from (9-6),

$$h_1^2 = \ell_{11}^2 + \ell_{12}^2 = 4^2 + 1^2 = 17$$

and the variance of X_1 can be decomposed as

$$\sigma_{11} = (\ell_{11}^2 + \ell_{12}^2) + \psi_1 = h_1^2 + \psi_1$$

or

$$\underbrace{19}_{\text{variance}} = \underbrace{4^2 + 1^2}_{\text{communality}} + \underbrace{2}_{\substack{\text{specific}\\\text{variance}}} = 17 + 2$$

A similar breakdown occurs for the other variables. ■

1.3 Invarianza rispetto a trasformazioni di scala

Assumiamo il modello fattoriale per x:

$$\underset{p\times1}{x} = \underset{p\times k}{\Lambda}\underset{k\times1}{f} + \underset{p\times1}{u}$$

Consideriamo una trasformazione di scala per x:

$$\underset{p\times1}{y} = \underset{p\times p}{A}\underset{p\times1}{x}$$

dove $\underset{p\times p}{A} = diag(a_1, \ldots, a_p)$ è una trasformazione di scala. Il modello fattoriale è ancora valido per y? Abbiamo

$$\begin{aligned}
\underset{p\times1}{y} &= \underset{p\times p}{A}\underset{p\times1}{x} \\
&= \underset{p\times p}{A}(\underset{p\times k}{\Lambda}\underset{k\times1}{f} + \underset{p\times1}{u}) \\
&= \underset{p\times p}{A}\underset{p\times k}{\Lambda}\underset{k\times1}{f} + \underset{p\times p}{A}\underset{p\times1}{u} \\
&= \underset{p\times k}{\Lambda_y}\underset{k\times1}{f} + \underset{p\times1}{u_y}
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{C}\text{ov}(y) &= \mathbb{C}\text{ov}(Ax) = A\mathbb{C}\text{ov}(x)A' = A\Sigma A' \\
&= A\Lambda\Lambda' A' + A\Psi A' \\
&= \Lambda_y\Lambda_y' + \Psi_y
\end{aligned}$$

quindi il modello fattoriale è ancora valido per y con pesi fattoriali $\Lambda_y = A\Lambda$ e varianze specifiche $\Psi_y = A\Psi A'$.
Il risultato precedente mostra che il modello fattoriale rimane essenzialmente inalterato se effettuiamo una trasformazione di scala. La standardizzazione

$$\underset{p\times 1}{z} = \underset{p\times p}{D}^{-1/2} \underset{p\times 1}{x}$$

è un caso particolare di trasformazione di scala dove

$$D^{-1/2} = diag(1/\sqrt{\sigma_{11}}, \ldots, 1/\sqrt{\sigma_{pp}})$$

Questo significa che, invece di considerare la decomposizione della matrice di varianze/covarianze di x, $\mathbb{C}\text{ov}(x)$, possiamo considerare la decomposizione della matrice di correlazione di x, $\mathbb{C}\text{orr}(x)$, o equivalentemente, la decomposizione della matrice di varianze/covarianze di z, $\mathbb{C}\text{ov}(z) = D^{-1/2}\Sigma D^{-1/2} = \mathbb{C}\text{orr}(x)$.
Si noti che sebbene il modello fattoriale è invariante rispetto a trasformazioni di scala, la *stima* dei parametri potrebbe essere influenzata dalle trasformazioni di scala.

1.4 Non-unicità dei pesi fattoriali

Sia $\underset{k\times k}{A}$ una matrice ortogonale: $AA' = A'A = I$

$$\begin{aligned}
\underset{p\times 1}{x} &= \underset{p\times k}{\Lambda}\underset{k\times 1}{f} + \underset{p\times 1}{u} \\
&= \underset{p\times k}{\Lambda}\underset{k\times k}{A}\underset{k\times k}{A'}\underset{k\times 1}{f} + \underset{p\times 1}{u} \\
&= \underset{p\times k}{\Lambda^*}\underset{k\times 1}{f^*} + \underset{p\times 1}{u}
\end{aligned}$$

- $\underset{p\times k}{\Lambda^*} = \underset{p\times k}{\Lambda}\underset{k\times k}{A}$
- $\underset{k\times 1}{f^*} = \underset{k\times k}{A'}\underset{k\times 1}{f}$
- $\mathbb{E}(f^*) = A'\mathbb{E}(f) = \underset{k\times 1}{0}$
- $\mathbb{C}\text{ov}(f^*) = A'\mathbb{C}\text{ov}(f)A = \underset{p\times p}{I}$
- $\mathbb{C}\text{ov}(x) = \Sigma = \Lambda\Lambda' + \Psi = \Lambda AA'\Lambda' + \Psi = \Lambda^*\Lambda^{*'} + \Psi$

Il risultato precedente mostra che il modello fattoriale con fattori comuni $\underset{k\times 1}{f}$ e pesi fattoriali $\underset{p\times k}{\Lambda}$, e il modello fattoriale con fattori comuni $\underset{k\times 1}{f^*}$ e pesi fattoriali $\underset{p\times k}{\Lambda^*}$ sono equivalenti per spiegare la matrice di varianza/covarianza Σ di $\underset{p\times 1}{x}$.

2 Stima del modello fattoriale

Obiettivo: determinare due matrici $\hat{\Lambda}$ e $\hat{\Psi}$ tali che $\widehat{\mathbb{C}\mathrm{ov}(x)} = \hat{\Sigma} = S = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$, oppure determinare due matrici $\hat{\Lambda}$ e $\hat{\Psi}$ tali che $\widehat{\mathbb{C}\mathrm{orr}(x)} = R = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$

2.1 Stima naïve

Example 2.1. *Si consideri il seguente esempio: sulla base di un campione di voti di studenti su tre materie,* x_1 (Classics)*,* x_2 (French) *e* x_3 (English) *si è ottenuta la seguente matrice di correlazione:*

$$\mathbf{R} = \begin{array}{l} \text{Classics} \\ \text{French} \\ \text{English} \end{array} \begin{pmatrix} 1.00 & & \\ 0.83 & 1.00 & \\ 0.78 & 0.67 & 1.00 \end{pmatrix}.$$

Si consideri il modello fattoriale con $k = 1$ *fattore*

$$\begin{aligned} x_1 &= \lambda_1 f + u_1 \\ x_2 &= \lambda_2 f + u_2 \\ x_3 &= \lambda_3 f + u_3 \end{aligned}$$

Le sei equazioni derivanti dall'uguaglianza $R = \Lambda\Lambda' + \Psi$ *sono*

$$\mathbf{R} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \begin{pmatrix} \lambda_1 & \lambda_2 \lambda_3 \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{pmatrix},$$

are

$$\begin{aligned} \hat{\lambda}_1 \lambda_2 &= 0.83, \\ \hat{\lambda}_1 \lambda_3 &= 0.78, \\ \hat{\lambda}_1 \lambda_4 &= 0.67, \\ \psi_1 &= 1.0 - \hat{\lambda}_1^2, \\ \psi_2 &= 1.0 - \hat{\lambda}_2^2, \\ \psi_3 &= 1.0 - \hat{\lambda}_3^2. \end{aligned}$$

The solutions of these equations are

$$\begin{aligned} \hat{\lambda}_1 = 0.99, \quad \hat{\lambda}_2 = 0.84, \quad \hat{\lambda}_3 = 0.79, \\ \hat{\psi}_1 = 0.02, \quad \hat{\psi}_2 = 0.30, \quad \hat{\psi}_3 = 0.38. \end{aligned}$$

Example 2.2. *Casi di Heywood*

Esempio 4 (tratto da Everitt e Dunn, 2001):

Si stimino i parametri del modello fattoriale ad un fattore per i punteggi ottenuti nelle tre materie:

x_1: Classics, x_2: French e x_3: English.

$$x_1 = \lambda_1 f + u_1,$$
$$x_2 = \lambda_2 f + u_2,$$
$$x_3 = \lambda_3 f + u_3.$$

data la matrice di correlazione campionaria
$$\mathbf{R} = \begin{array}{l} \\ \text{Classics} \\ \text{French} \\ \text{English} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Classics} & \text{French} & \text{English} \end{array} \\ \begin{pmatrix} 1.00 & & \\ 0.84 & 1.00 & \\ 0.60 & 0.35 & 1.00 \end{pmatrix} \end{array}.$$

Si ottiene il sistema
$$\hat{\psi}_1 = 1.0 - \hat{\lambda}_1^2, \quad \hat{\psi}_2 = 1.0 - \hat{\lambda}_2^2, \quad \hat{\psi}_3 = 1.0 - \hat{\lambda}_3^2.$$
e
$$\lambda_1\lambda_2 = 0.84, \quad \lambda_3\lambda_2 = 0.35, \quad \lambda_1\lambda_3 = 0.6$$

da cui la soluzione
$$\hat{\lambda}_1 = 1.2, \quad \hat{\lambda}_2 = 0.7, \quad \hat{\lambda}_3 = 0.5,$$
$$\hat{\psi}_1 = -0.44, \quad \hat{\psi}_2 = 0.51, \quad \hat{\psi}_3 = 0.75$$

Stime di tipo intuitivo possono condurre a soluzioni non accettabili anche se il modello è identificato esattamente!

Example 2.3. *Modello ad un fattore:* $\mathbb{C}\mathrm{orr}(x)$

Esempio 3 (tratto da Hardle e Simar 2003)

Si stimi il modello fattoriale ad un fattore per la matrice di correlazione relativa a valutazioni fornite da 40 clienti su 25 tipologie di auto per le variabili X_1: economicità, X_2: accessori, X_3: deprezzamento

$$\mathcal{R} = \begin{pmatrix} 1 & 0.975 & 0.613 \\ 0.975 & 1 & 0.620 \\ 0.613 & 0.620 & 1 \end{pmatrix}$$

Soluzione

$$\begin{pmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} \\ r_{X_1X_2} & 1 & r_{X_2X_3} \\ r_{X_1X_3} & r_{X_2X_3} & 1 \end{pmatrix} = \mathcal{R} = \begin{pmatrix} \hat{\lambda}_1^2 & \hat{\lambda}_1\hat{\lambda}_2 & \hat{\lambda}_1\hat{\lambda}_3 \\ & \hat{\lambda}_2^2 & \hat{\lambda}_2\hat{\lambda}_3 \\ & & \hat{\lambda}_3^2 \end{pmatrix} + \begin{pmatrix} \hat{\psi}_1 & 0 & 0 \\ & \hat{\psi}_2 & 0 \\ & & \hat{\psi}_3 \end{pmatrix}$$

Da cui

$$\frac{r_{X_1X_2} r_{X_1X_3}}{r_{X_2X_3}} = (0.982)^2 = \hat{\lambda}_1^2, \qquad \frac{r_{X_1X_2} r_{X_2X_3}}{r_{X_1X_3}} = (0.993)^2 = \hat{\lambda}_2^2 \qquad \frac{r_{X_1X_3} r_{X_2X_3}}{r_{X_1X_2}} = (0.624)^2 = \hat{\lambda}_3^2$$

e

$\psi_i = 1 - \hat{\lambda}_1^2 = 0.035$ $\psi_2 = 0.014$ $\psi_3 = 0.610$

Le prime due comunalità sono ≈ 1. X_1 e X_2 sono ben spiegate dal I° fattore (economicità + accessori)

Example 2.4. *Modello ad un fattore:* $\mathbb{C}\mathrm{ov}(x)$

Example 11.1 Let $p = 3$ and $k = 1$, then $d = 0$ and

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} q_1^2 + \psi_{11} & q_1 q_2 & q_1 q_3 \\ q_1 q_2 & q_2^2 + \psi_{22} & q_2 q_3 \\ q_1 q_3 & q_2 q_3 & q_3^2 + \psi_{33} \end{pmatrix}$$

with $\mathcal{Q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$ and $\Psi = \begin{pmatrix} \psi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & \psi_{33} \end{pmatrix}$. Note that here the constraint (11.8) is automatically verified since $k = 1$. We have

$$q_1^2 = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}; \qquad q_2^2 = \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}}; \qquad q_3^2 = \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}}$$

and

$$\psi_{11} = \sigma_{11} - q_1^2; \qquad \psi_{22} = \sigma_{22} - q_2^2; \qquad \psi_{33} = \sigma_{33} - q_3^2.$$

In this particular case ($k = 1$), the only rotation is defined by $\mathcal{G} = -1$, so the other solution for the loadings is provided by $-\mathcal{Q}$.

2.2 Vincoli e gradi di libertà

Numero di parametri del modello fattoriale $\Lambda\Lambda' + \Psi$: $pk + p$
Introduciamo ora il seguente vincolo (Vincolo 1):

$$\Lambda'\Psi^{-1}\Lambda = diag(b_1, \ldots, b_k)$$

con $b_1 \geq \ldots \geq b_k$
Il Vincolo 1 impone $k(k-1)/2$ restrizioni
Numero di parametri del modello fattoriale $\Lambda\Lambda' + \Psi$ dato il Vincolo 1: $pk + p - k(k-1)/2$
Come alternativa al Vincolo 1 si può considerare
Vincolo 2: $\Lambda' D^{-1}\Lambda = diag(c_1, \ldots, c_k)$ con $c_1 \geq \ldots \geq c_k$ e $D = diag(\sigma_{11}, \ldots, \sigma_{pp})$

I *gradi di libertà* (= numero dei parametri "liberi") sono dati dalla differenza tra i $p(p+1)/2$ parametri di $\underset{p\times p}{\Sigma}$ e il numero di parametri del modello fattoriale dato il Vincolo 1:

$$d = p(p+1)/2 - (pk + p - k(k-1)/2) = (p-k)^2/2 - (p+k)/2$$

- Se $d < 0$, allora il modello è indeterminato (ci sono infinite soluzioni)
- Se $d = 0$, allora la soluzione è unica (ma non necessariamente propria)
- $d > 0$, allora ci sono più equazioni che parametri: non c'è una soluzione esatta (ci si accontenta di una approssimazione)

Example 2.5. *Modello indeterminato*

Example 11.2 Suppose now $p = 2$ and $k = 1$, then $d < 0$ and

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} q_1^2 + \psi_{11} & q_1 q_2 \\ q_1 q_2 & q_2^2 + \psi_{22} \end{pmatrix}.$$

We have infinitely many solutions: for any α ($\rho < \alpha < 1$), a solution is provided by

$$q_1 = \alpha; \qquad q_2 = \rho/\alpha; \qquad \psi_{11} = 1 - \alpha^2; \qquad \psi_{22} = 1 - (\rho/\alpha)^2.$$

2.3 Metodi di stima

Data $\underset{p\times p}{\hat{\Sigma}} = \underset{p\times p}{S}$ (oppure $= \underset{p\times p}{R}$), vogliamo stimare $\underset{p\times p}{\hat{\Psi}}$ e $\underset{p\times k}{\hat{\Lambda}}$ in modo tale che $\hat{\Sigma} \approx \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$ e sia rispettato il Vincolo 1 o 2

- Naïve (senza vincolo)
- *Componenti principali*
- *Fattori principali*
- *Massima Verosimiglianza* (richiede assunzione di Normalità per $\underset{p\times 1}{x}$)

Rotazione dei fattori:
Dopo aver stimato il modello fattoriale, può essere utile ruotare i pesi fattoriali $\hat{\Lambda}$ per ottenere $\hat{\Lambda}^* = \hat{\Lambda}A$ (con A matrice ortogonale), al fine di trovare configurazioni più facilmente interpretabili

Numero di fattori:
In pratica, dobbiamo anche determinare il valore di k

2.4 Metodo dei fattori principali

Si parte da $\underset{p\times p}{\hat{\Sigma}} = \widehat{\mathbb{C}\mathrm{orr}(x)} = \underset{p\times p}{R}$ per trovare $\underset{p\times p}{\hat{\Psi}}$ e $\underset{p\times k}{\hat{\Lambda}}$ in modo tale che $R - \hat{\Psi} \approx \hat{\Lambda}\hat{\Lambda}'$ e sia rispettato il Vincolo 2

- $R^* = R - \hat{\Psi}$ è detta *matrice di correlazione ridotta*
- $\{\mathbb{C}\mathrm{orr}(x)\}_{ii} = 1 = h_i^2 + \psi_i$, quindi se abbiamo a disposizione una stima iniziale $\hat{h}_i^2$, allora $\{R^*\}_{ii} = 1 - \hat{\psi}_i = \hat{h}_i^2$
- $R^* = R - \hat{\Psi}$ è una matrice simmetrica, quindi la sua decomposizione spettrale è $R^* = VLV'$ con $L = diag(l_1, \dots, l_p)$ e $V = [v_1, \dots, v_p]$. Se i primi k autovalori $l_1, \dots, l_k$ sono positivi e i rimanenti $p - k$ autovalori $l_{k+1}, \dots, l_p$ prossimi a 0, allora

  $$R^* \approx V_k L_k V_k'$$

  dove $\underset{p\times k}{V_k}$ contiene le prime k colonne di V e $\underset{k\times k}{L_k} = diag(l_1, \dots, l_k)$

- Segue $R^* = R - \hat{\Psi} \approx (V_k L_k^{1/2})(V_k L_k^{1/2})' \approx \hat{\Lambda}\hat{\Lambda}'$, quindi
$$\hat{\Lambda} \approx V_k L_k^{1/2}$$

Inizializzazione:

- Partire dalla stima R della matrice di correlazione $\mathbb{C}\mathrm{orr}(x)$
- Calcolare la stima iniziale $\hat{h}_i^2$ della comunalità h_i^2 come
  - $\hat{h}_i^2 = \max_{j \neq i} |\widehat{\mathbb{C}\mathrm{orr}}(x_i, x_j)|$
  - $\hat{h}_i^2 = 1 - \frac{1}{r^{ii}}$ dove $r^{ii} = \{R^{-1}\}_{ii}$, che equivale il coefficiente di determinazione lineare multiplo tra x_i e $\underset{(p-1)\times 1}{x_{-i}}$
- Ottenere la matrice di correlazione ridotta R^* da R ma sostituendo i valori 1 sulla diagonale con $\hat{h}_1^2, \ldots, \hat{h}_p^2$

Algoritmo iterativo:

1. $R^* \leftarrow R$ e poi $\{R^*\}_{ii} \leftarrow \hat{h}_i^2$, $i = 1, \ldots, p$
2. Ottenere la decomposizione spettrale $R^* = VLV'$
3. Fissare k e determinare V_k e L_k
4. Stimare Λ con $\hat{\Lambda} \approx V_k L_k^{1/2}$
5. Aggiornare $\hat{h}_i^2 \leftarrow \sum_{j=1}^k \hat{\lambda}_{ij}^2$ e $\{R^*\}_{ii} \leftarrow \hat{h}_i^2$
6. Ripetere i passi 2-5 fino a raggiungere convergenza

Output: $\hat{\Lambda}$, $\hat{h}_i^2$ e $\hat{\psi}_i = 1 - \hat{h}_i^2$, $i = 1, \ldots, p$

Vincolo 2
$D = diag(\sigma_{11}, \ldots, \sigma_{pp}) = I$ perchè consideriamo la matrice di correlazione
Vincolo 2: $\hat{\Lambda}'\hat{D}^{-1}\hat{\Lambda} = \hat{\Lambda}'\hat{\Lambda} = diag(c_1, \ldots, c_k)$ con $c_1 \geq \ldots \geq c_k$
Quindi $\hat{\Lambda}$ soddisfa il Vincolo 2 perchè
$$\hat{\Lambda}'\hat{\Lambda} = (V_k L_k^{1/2})'(V_k L_k^{1/2}) = L_k = diag(l_1, \ldots, l_k)$$

Casi di Heywood
Nella procedura di stima iterativa possono succedere casi di Heywood, ovvero $\hat{\psi}_i < 0$ oppure $\hat{\psi}_i > 1$
$\hat{\psi}_i < 0$ non ha senso perchè ψ_i è una varianza, e quindi >0
$\hat{\psi}_i > 1$ non ha senso perchè $\mathbb{V}\mathrm{ar}(x_i) = 1$ è quindi $\psi_i \leq 1$

2.5 Stima di massima verosimiglianza

Assunzione: $\underset{p\times 1}{x}$ segue una distribuzione Normale p-variata $\mathcal{N}(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma})$

Funzione di log-verosimiglianza:

$$\begin{aligned}\ell(X;\mu,\Sigma) &= -\frac{1}{2}n\log|2\pi\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)\Sigma^{-1}(x_i-\mu)' \\ &= -\frac{1}{2}n\log|2\pi\Sigma| - \frac{1}{2}n\mathrm{tr}(\Sigma^{-1}S) - \frac{1}{2}n(\bar{x}-\mu)\Sigma^{-1}(\bar{x}-\mu)'\end{aligned}$$

Sostituendo $\hat{\mu} = \bar{x}$

$$\ell(X;\hat{\mu},\Sigma) = -\frac{n}{2}\Big\{\log|2\pi\Sigma| + \mathrm{tr}(\Sigma^{-1}S)\Big\}$$

e per $\Sigma = \Lambda\Lambda' + \Psi$ otteniamo

$$\ell(X;\hat{\mu},\Lambda,\Psi) = -\frac{n}{2}\Big\{\log|2\pi(\Lambda\Lambda'+\Psi)| + \mathrm{tr}[(\Lambda\Lambda'+\Psi)^{-1}S]\Big\}$$

Massimizzare

$$\ell(X;\hat{\mu},\Lambda,\Psi) = -\frac{n}{2}\Big\{\log|2\pi(\Lambda\Lambda'+\Psi)| + \mathrm{tr}[(\Lambda\Lambda'+\Psi)^{-1}S]\Big\}$$

rispetto a Ψ e Λ

Stima iterativa:

1. Per Ψ fissato, massimizza numericamente per Λ
2. Per Λ fissato, massimizza numericamente per Ψ

- Implementata nella funzione R `factanal()`
- Possiamo ottenere casi di Heywood

Example 2.6. *Voto di $n = 202$ studenti maschi su $p = 6$ esami (variabili) Gaelic (non-math), English (non-math), History (non-math), Arithmetic (math), Algebra (math), Geometry (math).*

$$\begin{array}{cc} & \begin{array}{cccccc}\text{Gaelic} & \text{English} & \text{History} & \text{Arithmetic} & \text{Algebra} & \text{Geometry}\end{array} \\ \mathbf{R} = & \begin{bmatrix} 1.0 & .439 & .410 & .288 & .329 & .248 \\ & 1.0 & .351 & .354 & .320 & .329 \\ & & 1.0 & .164 & .190 & .181 \\ & & & 1.0 & .595 & .470 \\ & & & & 1.0 & .464 \\ & & & & & 1.0 \end{bmatrix}\end{array}$$

Stima di massima verosimiglianza con $k = 2$

Table 9.5			
Variable	Estimated factor loadings F_1	F_2	Communalities $\hat{h}_i^2$
1. Gaelic	.553	.429	.490
2. English	.568	.288	.406
3. History	.392	.450	.356
4. Arithmetic	.740	−.273	.623
5. Algebra	.724	−.211	.569
6. Geometry	.595	−.132	.372

Stima di MV: $\hat{h}_1^2 = \hat{\lambda}_{11}^2 + \hat{\lambda}_{12}^2 = (0.553)^2 + (0.429)^2 \approx 0.490$
Primo fattore: intelligenza generale
Secondo fattore: abilità matematica vs abilità verbale

2.6 Rotazione dei pesi fattoriali

Per la rotazione dei pesi fattoriali $\underset{p\times k}{\Lambda}$, dobbiamo cercare una matrice ortogonale $\underset{k\times k}{A}$ ($A'A = AA' = I$) tale per cui i pesi fattoriali ruotati $\underset{p\times k}{\Lambda}^* = \underset{p\times k}{\Lambda}\underset{k\times k}{A}$ sono più facilmente interpretabili:

$$\underset{2\times 2}{A} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \text{ rotazione oraria per } k = 2$$

Questo non cambia la soluzione del modello, solo la sua descrizione. Situazione desiderata per i fini interpretativi:

- i pesi fattoriali sono tutti grandi e positivi o prossimi a 0 (con pochi valori intermedi)
- ogni variabile osservabile è legata in modo pesante al più ad un solo fattore

Per $k > 2$ il metodo *varimax* identifica la rotazione massimizzando un'opportuna funzione dei pesi fattoriali ruotati che misura la variabilità dei pesi.

Example 2.7. *Riprendiamo l'esempio precedente:*

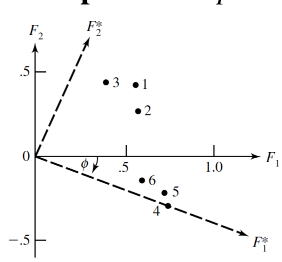

Figure 9.1 Factor rotation for test scores.

Table 9.6			
Variable	Estimated rotated factor loadings F_1^*	F_2^*	Communalities $\hat{h}_i^{*2} = \hat{h}_i^2$
1. Gaelic	.369	.594	.490
2. English	.433	.467	.406
3. History	.211	.558	.356
4. Arithmetic	.789	.001	.623
5. Algebra	.752	.054	.568
6. Geometry	.604	.083	.372

- *Primo fattore:* abilità matematica
- *Secondo fattore:* abilità verbale

2.7 Verifica d'ipotesi sul numero di fattori

Un vantaggio della stima di massima verosimiglianza e che permette un test di ipotesi sul numero di fattori
Ipotesi nulla H_0: il modello fattoriale con k fattori è corretto

$$\Sigma = \underset{p\times k}{\Lambda}\,\underset{k\times p}{\Lambda'} + \Psi$$

Ipotesi alternativa H_1: Σ è una matrice definitiva positiva diversa da quella specificata sotto l'ipotesi nulla
Rifiuto l'ipotesi nulla con un p-value $\leq 5\%$
Test sequenziali: parto da $k = 1$, se rifiuto proseguo con $k = 2, 3, \dots$ fino a quando fallisco di rifiutare l'ipotesi

Test rapporto di verosimiglianza

Siano $\hat{\Lambda}$ e $\hat{\Psi}$ le stime di massima verosimiglianza per il k specificato dall'ipotesi nulla
La statistica test rapporto di verosimiglianza è data da

$$T = -2\log\left(\frac{\text{MV sotto } H_0}{\text{MV}}\right) = n\log\left(\frac{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|}{|S|}\right)$$

e sotto H_0 segue asintoticamente una distribuzione

$$\chi^2_{\frac{1}{2}[(p-k)^2-p-k]}$$

Il p-value del test si calcola come $\Pr(\chi^2_{\frac{1}{2}[(p-k)^2-p-k]} > t)$ dove t è il valore osservato della statistica test
L'approssimazione χ^2 può essere migliorata utilizzando la statistica test con la correzione di Bartlett:

$$T_{Bartlett} = [(n-1) - (2p+4k+5)/6]\log\left(\frac{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|}{|S|}\right)$$

2.8 Stima dei punteggi fattoriali

I punteggi fattoriali $\underset{k\times 1}{\hat{f}} = (\hat{f}_1, \ldots, \hat{f}_k)'$ sono le "stime" delle variabili non osservabili $\underset{k\times 1}{f} = (f_1, \ldots, f_k)'$

Metodo di Thompson (1951)
La distribuzione condizionata di $\underset{k\times 1}{f}$ dato $\underset{p\times 1}{x}$ è

$$\mathcal{N}(\Lambda'\Sigma^{-1}x, I - \Lambda'\Psi^{-1}\Lambda)$$

Per l'i-sima osservazione x_i (se standardizzata z_i),

$$\hat{f}_i = \hat{\Lambda}'S^{-1}x_i \qquad (\hat{f}_i = \hat{\Lambda}'R^{-1}z_i)$$

Metodo di Bartlett (1937)

$$\hat{f}_i = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}x_i$$