

MAT-269: Sesión 16, Modelo de Análisis Factorial

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Idea:

El objetivo es intentar explicar la correlación entre un conjunto grande de variables en términos de un número pequeño de factores.

Este modelo supone que aquellos factores no son observables y son considerados variables aleatorias. Este modelo es muy apropiado en áreas como Psicología.

Definición 1 (Modelo de análisis factorial)

Sea $\mathbf{x} = (x_1, \dots, x_p)^\top$ vector aleatorio con $E(\mathbf{x}) = \boldsymbol{\mu}$ y $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$. Suponga que los elementos de \mathbf{x} satisfacen:

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\epsilon},$$

donde $\mathbf{z} = (z_1, \dots, z_m)^\top$ es llamado **factores comunes** ($m < p$) y $\boldsymbol{\Gamma} = (\gamma_{jk}) \in \mathbb{R}^{p \times m}$, con γ_{jk} llamado la **carga factorial** (loadings) de la variable j (x_j) sobre el factor z_k .



El modelo está basado en los siguientes supuestos:

$$E(\epsilon) = \mathbf{0}, \quad \text{Cov}(\epsilon) = \Psi = \text{diag}(\psi_1^2, \dots, \psi_p^2),$$

independientes de \mathbf{z} con

$$E(\mathbf{z}) = \mathbf{0}, \quad \text{Cov}(\mathbf{z}) = \mathbf{I}.$$

De este modo,

$$E(\mathbf{x}) = \boldsymbol{\mu} + \mathbf{\Gamma} E(\mathbf{z}) + E(\epsilon) = \boldsymbol{\mu},$$

y

$$\begin{aligned} \text{Cov}(\mathbf{x}) &= \text{Cov}(\mathbf{\Gamma} \mathbf{z} + \epsilon) = \mathbf{\Gamma} \text{Cov}(\mathbf{z}) \mathbf{\Gamma}^\top + \text{Cov}(\epsilon) \\ &= \mathbf{\Gamma} \mathbf{\Gamma}^\top + \Psi = \Sigma. \end{aligned}$$



Ejemplo: (Spearman, 1904)

Se desea examinar el desempeño de un grupo de niños en **Classics** (x_1), **French** (x_2) y **English** (x_3). Se calculó la matriz

$$\Sigma = \begin{pmatrix} 1.00 & 0.83 & 0.78 \\ 0.83 & 1.00 & 0.67 \\ 0.78 & 0.67 & 1.00 \end{pmatrix},$$

se puede verificar que $\Sigma = \Gamma\Gamma^\top + \Sigma$ con

$$\Gamma = \begin{pmatrix} 0.983 \\ 0.844 \\ 0.744 \end{pmatrix}, \quad \Psi = \begin{pmatrix} 0.34 & 0.00 & 0.00 \\ 0.00 & 0.29 & 0.00 \\ 0.00 & 0.00 & 0.37 \end{pmatrix},$$

luego el modelo asume la forma

$$\mathbf{x} = \boldsymbol{\mu} + \Gamma\mathbf{z} + \boldsymbol{\epsilon},$$

es decir

$$x_1 = \mu_1 + \gamma_1 z + \epsilon_1, \quad x_2 = \mu_2 + \gamma_2 z + \epsilon_2, \quad x_3 = \mu_3 + \gamma_3 z + \epsilon_3,$$

con z el factor “**inteligencia**”.



Observación:

Sea P matriz ortogonal $m \times m$, entonces,

$$\begin{aligned}x &= \mu + \Gamma z + \epsilon = \mu + (\Gamma P)(P^\top z) + \epsilon \\ &= \mu + \Gamma_* z_* + \epsilon,\end{aligned}$$

donde

$$\text{Cov}(z_*) = P^\top \text{Cov}(z) P = P^\top P = I,$$

de este modo el modelo **no es único**, es decir, existe un problema de **identificabilidad**.

El problema de no unicidad puede ser resuelto mediante rotar las cargas factoriales de tal manera que se satisfaga

$$\Gamma^\top \Psi^{-1} \Gamma \quad \text{sea diagonal,}$$

esto incorpora $m(m-1)/2$ restricciones.



Note que los elementos diagonales de $\Sigma (= \Gamma\Gamma^\top + \Psi)$ asumen la forma

$$\sigma_{ii} = \sum_{k=1}^m \gamma_{ik}^2 + \psi_i^2 = h_i^2 + \psi_i^2,$$

es decir, podemos descomponer la varianza de x_i en dos partes h_i^2 llamada **comunalidad** o varianza común y ψ_i^2 es llamada varianza residual.



Suponga una muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$ muestra aleatoria desde $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces

$$\ell(\boldsymbol{\theta}) = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}),$$

es conveniente perfilar la verosimilitud en $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$, obteniendo

$$\ell(\boldsymbol{\Gamma}, \boldsymbol{\Psi}) = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi}| - \frac{1}{2} \text{tr}(\boldsymbol{\Gamma}^\top \boldsymbol{\Gamma} + \boldsymbol{\Psi})^{-1} \mathbf{Q},$$

donde $\mathbf{Q} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$.



Se debe optimizar $\ell(\mathbf{\Gamma}, \mathbf{\Psi})$ sujeto a la condición que $\mathbf{\Gamma}^\top \mathbf{\Psi}^{-1} \mathbf{\Gamma} = \mathbf{\Delta}$, $\mathbf{\Delta}$ diagonal. Tomando derivadas con respecto a $\mathbf{\Gamma}$ y $\mathbf{\Psi}$ lleva a las siguientes ecuaciones:

$$\mathbf{S} \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}} (\mathbf{I} + \hat{\mathbf{\Gamma}}^\top \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{\Gamma}})$$

$$\hat{\mathbf{\Psi}} = \text{diag}(\mathbf{S} - \hat{\mathbf{\Gamma}} \hat{\mathbf{\Gamma}}^\top)$$

$$\hat{\mathbf{\Delta}} = \hat{\mathbf{\Gamma}}^\top \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{\Gamma}}.$$

Estas ecuaciones deben ser resueltas de forma iterativa y es bien conocido que sufre de problemas de convergencia.

Observación:

La función de log-verosimilitud puede no tener máximo, sujeto a $\mathbf{\Psi} > \mathbf{0}$, frecuentemente esto se debe a que uno o más elementos de $\mathbf{\Psi}$ tienden a cero.

Hasta que surgiera el algoritmo propuesto por Jöreskog (1967) y Lawley (1967) existía mucha dificultad en la estimación para el modelo factorial.¹

¹Estas rutinas se encuentran implementadas en el software LISREL.



Estimación ML: Método de componentes principales

Note que $\Sigma - \Psi = \Gamma\Gamma^\top$ es semidefinida positiva de rango m . De ahí que existe una matriz ortogonal M tal que

$$M^\top(\Sigma - \Psi)M = \text{diag}(\phi_1, \dots, \phi_m, 0, \dots, 0) = \Phi,$$

donde ϕ_j son los valores propios positivos. Sea M_1 la matriz que contiene los primeros m vectores propios y $\Phi_1 = \text{diag}(\phi_1, \dots, \phi_m)$. Entonces

$$\begin{aligned}\Sigma - \Psi &= M\Phi M^\top = (M_1, M_2) \begin{pmatrix} \Phi_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} M_1^\top \\ M_2^\top \end{pmatrix} \\ &= M_1\Phi_1M_1^\top = (\Gamma_1\Phi_1^{1/2})(\Gamma_1\Phi_1^{1/2})^\top,\end{aligned}$$

y $\Gamma = \Gamma_1\Phi_1^{1/2}$ es una solución. De este modo, para un estimador dado $\hat{\Psi}_1$ de Ψ se obtiene las primeras m **componentes principales** de $\hat{\Sigma} - \hat{\Psi}_1$ lo que permite estimar Γ como $\hat{\Gamma}_1\hat{\Phi}_1^{1/2}$.



Considere el modelo de análisis factorial,

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma} \mathbf{z}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n,$$

asumiremos que $\mathbf{z}_1, \dots, \mathbf{z}_n$ son vectores **no observables (missing)** IID $N_m(\mathbf{0}, \mathbf{I})$, independiente de los errores $\boldsymbol{\epsilon}_i \stackrel{\text{IID}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Psi})$, con $\boldsymbol{\Psi} = \text{diag}(\psi_1^2, \dots, \psi_p^2)$.

De este modo, condicional a los \mathbf{z}_i , tenemos:

$$\mathbf{x}_i | \mathbf{z}_i \stackrel{\text{IND}}{\sim} N_p(\boldsymbol{\mu} + \boldsymbol{\Gamma} \mathbf{z}_i, \boldsymbol{\Psi}), \quad \mathbf{z}_i \stackrel{\text{IND}}{\sim} N_m(\mathbf{0}, \mathbf{I}).$$

Integrando con relación a \mathbf{z}_i obtenemos la distribución marginal

$$\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\top + \boldsymbol{\Psi}), \quad i = 1, \dots, n.$$



De este modo, el vector de **datos completos** (\mathbf{x}, \mathbf{z}) con $\mathbf{z} = (z_1^\top, \dots, z_n^\top)^\top$ el **vector de datos perdidos**, podemos notar que

$$\mathbf{z}_i | \mathbf{x}_i \sim \mathcal{N}(\mathbf{\Gamma}^\top \mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}), (\mathbf{\Gamma}^\top \mathbf{\Sigma}^{-1} \mathbf{\Gamma} + \mathbf{I})^{-1}).$$

En este caso tenemos que la **log-verosimilitud de datos completos** asume la forma

$$\begin{aligned} \ell_c(\boldsymbol{\theta}; \mathbf{x}, \mathbf{z}) = & -\frac{n}{2} \log |2\pi \mathbf{\Psi}| \\ & - \frac{1}{2} \sum_{i=1}^n \{(\mathbf{x}_i - \boldsymbol{\mu} - \mathbf{\Gamma} \mathbf{z}_i)^\top \mathbf{\Psi}^{-1}(\mathbf{x}_i - \boldsymbol{\mu} - \mathbf{\Gamma} \mathbf{z}_i) + \mathbf{z}_i^\top \mathbf{z}_i\}. \end{aligned}$$

Note que,

$$\mathbb{E}(\mathbf{z}_i | \mathbf{x}_i) = (\mathbf{\Gamma}^\top \mathbf{\Sigma}^{-1} \mathbf{\Gamma} + \mathbf{I})^{-1} \mathbf{\Gamma}^\top \mathbf{\Psi}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}).$$



Substituyendo μ por \bar{x} y como:

$$\begin{aligned} (x_i - \bar{x} - \Gamma z_i)^\top \Psi^{-1} (x_i - \bar{x} - \Gamma z_i) &= (x_i - \bar{x})^\top \Psi^{-1} (x_i - \bar{x}) \\ &\quad + z_i^\top \Gamma^\top \Psi^{-1} \Gamma z_i - 2(x_i - \bar{x})^\top \Psi^{-1} \Gamma z_i, \end{aligned}$$

usando resultados básicos de esperanzas de formas cuadráticas, tenemos

$$\begin{aligned} E\{(x_i - \bar{x} - \Gamma z_i)^\top \Psi^{-1} (x_i - \bar{x} - \Gamma z_i) | x_i, \theta^{(k)}\} &= (x_i - \bar{x})^\top \Psi^{-1} (x_i - \bar{x}) \\ &\quad - 2(x_i - \bar{x})^\top \Psi^{-1} \Gamma \hat{z}_i + \text{tr} \Gamma^\top \Psi^{-1} \Gamma \text{Cov}(z_i | x_i, \theta^{(k)}) + \hat{z}_i^\top \Gamma^\top \Psi^{-1} \Gamma \hat{z}_i \\ &= (x_i - \bar{x} - \Gamma \hat{z}_i)^\top \Psi^{-1} (x_i - \bar{x} - \Gamma \hat{z}_i) + \text{tr} \Gamma^\top \Psi^{-1} \Gamma \hat{\Omega}, \end{aligned}$$

donde

$$\begin{aligned} \hat{z}_i &= E(z_i | x_i, \theta^{(k)}) = (\hat{\Gamma}^\top \hat{\Psi}^{-1} \hat{\Gamma} + I)^{-1} \hat{\Gamma}^\top \hat{\Psi}^{-1} (x_i - \bar{x}), \\ \hat{\Omega} &= \text{Cov}(z_i | x_i, \theta^{(k)}) = (\hat{\Gamma}^\top \hat{\Psi}^{-1} \hat{\Gamma} + I)^{-1}, \end{aligned}$$

con $\hat{\Gamma}$ y $\hat{\Psi}$ siendo evaluadas en $\theta^{(k)}$, y análogamente

$$E(z_i^\top z_i | x_i, \theta^{(k)}) = \hat{z}_i^\top \hat{z}_i + \text{tr} \hat{\Omega}.$$



De este modo,

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = -\frac{n}{2} \log |\boldsymbol{\Psi}| \\ - \frac{1}{2} \sum_{i=1}^n \left\{ (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i)^\top \boldsymbol{\Psi}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i) + \text{tr} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \right\}.$$

Tenemos que

$$\begin{aligned} d_{\boldsymbol{\Psi}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) &= -\frac{n}{2} \text{tr} \boldsymbol{\Psi}^{-1} + \frac{n}{2} \text{tr} \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi} \\ &\quad + \frac{1}{2} \text{tr} \boldsymbol{\Psi}^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i) (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i)^\top \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi} \\ &= \frac{1}{2} \text{tr} \boldsymbol{\Psi}^{-1} \{ \mathbf{S} + n \boldsymbol{\Gamma} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^\top - n \boldsymbol{\Psi} \} \boldsymbol{\Psi}^{-1} d \boldsymbol{\Psi}. \end{aligned}$$

De ahí que

$$\boldsymbol{\Psi}^{(k+1)} = \text{diag} \left(\frac{1}{n} \mathbf{S} + \boldsymbol{\Gamma}^{(k)} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^{(k)\top} \right)$$



Por otro lado,

$$\begin{aligned}d_{\Gamma} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i)^{\top} \boldsymbol{\Psi}^{-1} (d \boldsymbol{\Gamma}) \hat{\mathbf{z}}_i \\&\quad - \frac{n}{2} \text{tr} \hat{\boldsymbol{\Omega}} (d \boldsymbol{\Gamma})^{\top} \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} - \frac{n}{2} \text{tr} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^{\top} \boldsymbol{\Psi}^{-1} d \boldsymbol{\Gamma} \\&= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i)^{\top} \boldsymbol{\Psi}^{-1} (d \boldsymbol{\Gamma}) \hat{\mathbf{z}}_i - n \text{tr} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^{\top} \boldsymbol{\Psi}^{-1} d \boldsymbol{\Gamma},\end{aligned}$$

es decir,

$$d_{\Gamma} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = \text{tr} \sum_{i=1}^n \hat{\mathbf{z}}_i (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i)^{\top} \boldsymbol{\Psi}^{-1} d \boldsymbol{\Gamma} - n \text{tr} \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^{\top} \boldsymbol{\Psi}^{-1} d \boldsymbol{\Gamma}.$$

Desde la condición de primer orden, sigue que

$$\begin{aligned}\sum_{i=1}^n \hat{\mathbf{z}}_i (\mathbf{x}_i - \bar{\mathbf{x}} - \boldsymbol{\Gamma} \hat{\mathbf{z}}_i)^{\top} - n \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^{\top} &= \mathbf{0} \\ \sum_{i=1}^n \hat{\mathbf{z}}_i (\mathbf{x}_i - \bar{\mathbf{x}})^{\top} - \sum_{i=1}^n \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^{\top} \boldsymbol{\Gamma}^{\top} - n \hat{\boldsymbol{\Omega}} \boldsymbol{\Gamma}^{\top} &= \mathbf{0}\end{aligned}$$



Es decir,

$$n \left(\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\top + \hat{\mathbf{\Omega}} \right) \mathbf{\Gamma}^\top = \sum_{i=1}^n \hat{\mathbf{z}}_i (\mathbf{x}_i - \bar{\mathbf{x}})^\top.$$

Finalmente,

$$\mathbf{\Gamma}^{(k+1)\top} = \left(\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{z}}_i \hat{\mathbf{z}}_i^\top + \hat{\mathbf{\Omega}} \right)^{-1} \sum_{i=1}^n \hat{\mathbf{z}}_i (\mathbf{x}_i - \bar{\mathbf{x}})^\top.$$

Observaciones:

- ▶ Este algoritmo EM es simple y numéricamente estable.
- ▶ Posibilidad de soluciones múltiples (que no necesariamente son rotaciones unas de otras).
- ▶ Lenta razón de convergencia, lentitud que se atribuye a la proporción de datos perdidos.
- ▶ Errores estándar no son obtenidos como un subproducto del procedimiento de estimación (diferente a algoritmo Newton-Raphson).

