

# Análisis Factorial

Un estudio teórico con demostraciones completas

Basado en los apuntes de Norberto Corral Blanco y Beatriz Sinova Fernández

Grado en Matemáticas — Universidad de Oviedo

## Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. El Modelo Factorial</b>	<b>2</b>
2.1. Formulación . . . . .	2
2.2. Hipótesis básicas . . . . .	3
2.3. Necesidad de las hipótesis básicas . . . . .	3
<b>3. Consecuencias del Modelo: Estructura de Covarianzas</b>	<b>4</b>
3.1. Descomposición de la matriz de covarianzas . . . . .	4
3.2. Comunalidad y especificidad . . . . .	5
3.3. Covarianza entre variables y entre variables y factores . . . . .	5
<b>4. Unicidad del Modelo Factorial</b>	<b>6</b>
4.1. Fuentes de indeterminación . . . . .	6
4.2. Condiciones de unicidad: el criterio $\Lambda^t\Lambda$ diagonal . . . . .	7
<b>5. Número Máximo de Factores</b>	<b>7</b>
<b>6. Estimación: El Método del Factor Principal</b>	<b>8</b>
6.1. Planteamiento . . . . .	8
6.2. Algoritmo iterativo . . . . .	8
6.3. Algoritmo completo . . . . .	9
<b>7. Rotación de Factores</b>	<b>9</b>
7.1. Motivación . . . . .	9
7.2. Criterio varimax . . . . .	10
<b>8. Resumen del Procedimiento</b>	<b>10</b>
<b>A. Apéndice: Propiedades de la Raíz Cuadrada de una Matriz</b>	<b>10</b>

# 1. Introducción

El **análisis factorial** es una técnica estadística multivariante cuyo objetivo es explicar un conjunto de  $p$  variables observadas a través de un número reducido  $m < p$  de variables latentes, no observadas directamente, denominadas *factores*.

La idea central puede ilustrarse con el siguiente ejemplo motivador.

**Ejemplo 1.1** (Medidas corporales). Si tomamos muchas medidas físicas del cuerpo humano (estatura, longitud de brazos y piernas, anchura de hombros, peso, etc.), es intuitivo que no son independientes. Conocidas algunas de ellas, es posible prever las restantes con relativa precisión. Una explicación natural es que todas estas dimensiones dependen de ciertos *factores subyacentes* (por ejemplo, la constitución genética, el estado nutricional), que, si fuesen conocidos, permitirían predecir las variables observadas con un error pequeño.

Desde el punto de vista histórico, la técnica nació del interés de Karl Pearson y Charles Spearman en las décadas de 1900–1930 por comprender las dimensiones de la inteligencia humana. Gran parte de sus avances se produjeron en el ámbito de la *psicometría*.

**Objetivo formal:** Explicar por qué una serie de variables observadas están correladas entre sí mediante un número reducido de factores subyacentes, de comportamiento aleatorio y no observable, que están presentes en todas las variables.

**Observación 1.1.** La diferencia fundamental con las *componentes principales* es que éstas buscan explicar las **varianzas**, mientras que los factores se construyen para explicar las **covarianzas o correlaciones** entre las variables.

## 2. El Modelo Factorial

### 2.1. Formulación

#### Modelo Factorial

Sea  $\vec{x}$  un vector aleatorio de dimensión  $p \times 1$ . El **modelo factorial** con  $m$  factores ( $m < p$ ) establece que

$$\vec{x} = \vec{\mu} + \Lambda \vec{f} + \vec{u},$$

donde:

- $\vec{\mu} \in \mathbb{R}^p$ : vector de medias de  $\vec{x}$ .
- $\Lambda \in \mathbb{R}^{p \times m}$ : *matriz de pesos (cargas) factoriales*, que describe cómo afectan los factores a cada variable.
- $\vec{f} \in \mathbb{R}^m$ : vector de *factores comunes* (variables latentes).
- $\vec{u} \in \mathbb{R}^p$ : vector de *factores específicos* (perturbaciones idiosincráticas).

Componente a componente, el modelo se escribe como

$$x_i = \mu_i + \sum_{j=1}^m \lambda_{ij} f_j + u_i, \quad i = 1, \dots, p.$$

Los factores  $\vec{f}$  justifican la *asociación* que existe entre las variables  $x_i$ , mientras que  $\vec{u}$  recoge todo el efecto no compartido.

## 2.2. Hipótesis básicas

Las hipótesis del modelo son las siguientes:

**H1.**  $\mathbb{E}(\vec{f}) = \vec{0}$  (factores comunes centrados).

**H2.**  $\text{Cov}(\vec{f}) = I_m$  (factores comunes estandarizados e incorrelados).

**H3.**  $\mathbb{E}(\vec{u}) = \vec{0}$  (factores específicos centrados).

**H4.**  $\text{Cov}(\vec{u}) = \Psi = \text{diag}(\psi_{11}, \dots, \psi_{pp})$  (factores específicos incorrelados entre sí).

**H5.**  $\text{Cov}(\vec{f}, \vec{u}) = 0$  (independencia entre factores comunes y específicos).

**Observación 2.1.** En la bibliografía se añade a veces la hipótesis de normalidad conjunta de  $\vec{f}$  y  $\vec{u}$ , pero ésta no es estrictamente necesaria para la formulación del modelo ni para la estimación por el método del factor principal.

## 2.3. Necesidad de las hipótesis básicas

A continuación justificamos cuáles de las cinco hipótesis son *esenciales* y cuáles pueden relajarse sin pérdida de generalidad.

**Proposición 2.1.** *Las hipótesis H1 y H3 (centrado de  $\vec{f}$  y  $\vec{u}$ ) no son restrictivas: si no se verifican, se puede reformular el modelo de manera equivalente que sí las satisfaga.*

*Demostración. Respecto a H1.* Supongamos que  $\mathbb{E}(\vec{f}) = \vec{\mu}_f \neq \vec{0}$ . Entonces

$$\vec{x} = \vec{\mu} + \Lambda \vec{f} + \vec{u} = \vec{\mu} + \Lambda(\vec{f} - \vec{\mu}_f) + \Lambda \vec{\mu}_f + \vec{u} = \vec{\mu}^* + \Lambda \vec{f}^* + \vec{u},$$

donde  $\vec{\mu}^* = \vec{\mu} + \Lambda \vec{\mu}_f$  y  $\vec{f}^* = \vec{f} - \vec{\mu}_f$ . Como  $\mathbb{E}(\vec{f}^*) = \mathbb{E}(\vec{f}) - \vec{\mu}_f = \vec{0}$ , el nuevo modelo satisface H1.

**Respecto a H3.** El razonamiento es análogo. Si  $\mathbb{E}(\vec{u}) = \vec{\mu}_u \neq \vec{0}$ , definimos  $\vec{\mu}^* = \vec{\mu} + \mu_u$  y  $\vec{u}^* = \vec{u} - \mu_u$ . Entonces  $\mathbb{E}(\vec{u}^*) = \vec{0}$  y el modelo  $\vec{x} = \vec{\mu}^* + \Lambda \vec{f} + \vec{u}^*$  es equivalente.  $\square$

**Proposición 2.2.** *La hipótesis H2 ( $\text{Cov}(\vec{f}) = I_m$ ) no es restrictiva: si los factores tienen covarianzas arbitrarias, siempre existe una parametrización equivalente que satisface H2.*

*Demostración. Caso 1:  $\text{Cov}(\vec{f}) = D$  diagonal.* Sea  $D = \text{diag}(\sigma_{f_1}^2, \dots, \sigma_{f_m}^2)$  y  $D^{1/2} = \text{diag}(\sigma_{f_1}, \dots, \sigma_{f_m})$ . Entonces

$$\vec{x} = \vec{\mu} + \Lambda \vec{f} + \vec{u} = \vec{\mu} + \Lambda D^{1/2} D^{-1/2} \vec{f} + \vec{u} = \vec{\mu} + \Lambda^* \vec{f}^* + \vec{u},$$

con  $\Lambda^* = \Lambda D^{1/2}$  y  $\vec{f}^* = D^{-1/2} \vec{f}$ . Calculamos

$$\text{Cov}(\vec{f}^*) = D^{-1/2} \text{Cov}(\vec{f}) D^{-1/2} = D^{-1/2} D D^{-1/2} = I_m.$$

**Caso 2:**  $\text{Cov}(\vec{f}) = A$  **no diagonal con**  $|A| \neq 0$ . Como  $A$  es simétrica semidefinida positiva, admite la descomposición espectral  $A = UDU^t$ , donde  $U$  es ortogonal y  $D$

diagonal. Definimos  $A^{1/2} = UD^{1/2}U^t$  (raíz cuadrada simétrica de  $A$ ) y  $A^{-1/2} = UD^{-1/2}U^t$ . Entonces

$$\vec{x} = \vec{\mu} + \Lambda A^{1/2} A^{-1/2} \vec{f} + \vec{u} = \vec{\mu} + \Lambda^* \vec{f}^* + \vec{u},$$

con  $\Lambda^* = \Lambda A^{1/2}$  y  $\vec{f}^* = A^{-1/2} \vec{f}$ . Verificamos que  $A^{-1/2}$  es simétrica:

$$(A^{-1/2})^t = (UD^{-1/2}U^t)^t = UD^{-1/2}U^t = A^{-1/2},$$

y que la nueva covarianza es la identidad:

$$\text{Cov}(\vec{f}^*) = A^{-1/2} \text{Cov}(\vec{f})(A^{-1/2})^t = A^{-1/2} \Lambda A^{-1/2} = I_m.$$

**Caso 3:**  $|A| = 0$ . En este caso algún factor, digamos  $f_m$ , es combinación lineal de los demás:  $f_m = a_1 f_1 + \dots + a_{m-1} f_{m-1}$ . Sustituyendo en la expresión de  $x_i$ :

$$x_i = \mu_i + \sum_{j=1}^{m-1} \lambda_{ij} f_j + \lambda_{im} f_m + u_i = \mu_i + \sum_{j=1}^{m-1} (\lambda_{ij} + \lambda_{im} a_j) f_j + u_i,$$

lo que define un modelo equivalente con  $m-1$  factores y cargas  $\lambda_{ij}^* = \lambda_{ij} + \lambda_{im} a_j$ . Se puede continuar reduciendo factores hasta alcanzar una matriz de covarianzas no singular.  $\square$

**Observación 2.2.** La hipótesis H2 se impone para que la interpretación sea más sencilla y para garantizar la identificabilidad del modelo (sin ella, la misma covarianza observada  $\Sigma$  puede generarse con infinitas matrices  $\Lambda$  diferentes).

**Proposición 2.3.** *Las hipótesis H4 y H5 son genuinamente restrictivas y hacen parte esencial del modelo factorial.*

*Demostración. Respecto a H4.* Cada factor específico  $u_i$  afecta a una única variable  $x_i$  por construcción; es por tanto natural exigir que distintos factores específicos no estén relacionados entre sí. Si existiese correlación entre  $u_i$  y  $u_k$ , ésta podría absorberse en los factores comunes, perdiendo la distinción entre la parte compartida y la parte específica.

*Respecto a H5.* Si  $\text{Cov}(\vec{f}, \vec{u}) \neq 0$ , los factores específicos influirían en los factores comunes, dejando de ser específicos de una sola variable. Esto haría imposible separar la variabilidad compartida de la idiosincrática.  $\square$

## 3. Consecuencias del Modelo: Estructura de Covarianzas

### 3.1. Descomposición de la matriz de covarianzas

El resultado fundamental del análisis factorial es la expresión de  $\Sigma = \text{Cov}(\vec{x})$  en términos de los parámetros del modelo.

**Teorema 3.1** (Descomposición factorial de  $\Sigma$ ). *Bajo las hipótesis H1–H5,*

$$\Sigma = \text{Cov}(\vec{x}) = \Lambda \Lambda^t + \Psi.$$

*Demostración.* Como  $\mathbb{E}(\vec{x}) = \vec{\mu}$ , tenemos  $\vec{x} - \vec{\mu} = \Lambda \vec{f} + \vec{u}$ . Por tanto,

$$\begin{aligned}\Sigma &= \mathbb{E}[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^t] = \mathbb{E}[(\Lambda \vec{f} + \vec{u})(\Lambda \vec{f} + \vec{u})^t] \\ &= \mathbb{E}[\Lambda \vec{f} \vec{f}^t \Lambda^t] + \mathbb{E}[\Lambda \vec{f} \vec{u}^t] + \mathbb{E}[\vec{u} \vec{f}^t \Lambda^t] + \mathbb{E}[\vec{u} \vec{u}^t].\end{aligned}$$

Analizamos cada término:

$$\begin{aligned}\mathbb{E}[\Lambda \vec{f} \vec{f}^t \Lambda^t] &= \Lambda \text{Cov}(\vec{f}) \Lambda^t \stackrel{\text{H2}}{=} \Lambda I_m \Lambda^t = \Lambda \Lambda^t, \\ \mathbb{E}[\Lambda \vec{f} \vec{u}^t] &= \Lambda \text{Cov}(\vec{f}, \vec{u}) \stackrel{\text{H5}}{=} \Lambda \cdot 0 = 0, \\ \mathbb{E}[\vec{u} \vec{f}^t \Lambda^t] &= \text{Cov}(\vec{u}, \vec{f}) \Lambda^t \stackrel{\text{H5}}{=} 0, \\ \mathbb{E}[\vec{u} \vec{u}^t] &= \text{Cov}(\vec{u}) \stackrel{\text{H4}}{=} \Psi.\end{aligned}$$

Sumando los cuatro términos obtenemos  $\Sigma = \Lambda \Lambda^t + \Psi$ . □

### 3.2. Comunalidad y especificidad

Consideramos el elemento diagonal  $(i, i)$  de la descomposición anterior.

**Definición 3.1** (Comunalidad y especificidad). Para la variable  $x_i$ , la **varianza** se descompone como

$$\sigma_i^2 = \text{Var}(x_i) = \sum_{j=1}^m \lambda_{ij}^2 + \psi_{ii} = h_i^2 + \psi_{ii},$$

donde:

- $h_i^2 = \sum_{j=1}^m \lambda_{ij}^2$  es la **comunalidad** de  $x_i$ : fracción de la varianza de  $x_i$  explicada por los factores comunes.
- $\psi_{ii}$  es la **especificidad** (o unicidad) de  $x_i$ : fracción de la varianza de  $x_i$  no compartida con las demás variables.

*Demostración.* Aplicamos el Teorema 3.1 al elemento  $(i, i)$ :

$$\sigma_i^2 = (\Lambda \Lambda^t + \Psi)_{ii} = (\Lambda \Lambda^t)_{ii} + \psi_{ii}.$$

Calculamos  $(\Lambda \Lambda^t)_{ii}$ : la fila  $i$  de  $\Lambda$  es  $(\lambda_{i1}, \dots, \lambda_{im})$ , luego

$$(\Lambda \Lambda^t)_{ii} = \sum_{j=1}^m \lambda_{ij}^2 = h_i^2. \quad \square$$

### 3.3. Covarianza entre variables y entre variables y factores

**Proposición 3.1** (Covarianza entre variables). Para  $i \neq k$ ,

$$\text{Cov}(x_i, x_k) = \sum_{j=1}^m \lambda_{ij} \lambda_{kj} = (\Lambda \Lambda^t)_{ik}.$$

En particular, la asociación entre  $x_i$  y  $x_k$  depende exclusivamente de los factores comunes.

*Demostración.* Dado que  $x_i - \mu_i = \sum_{j=1}^m \lambda_{ij} f_j + u_i$  y  $x_k - \mu_k = \sum_{j=1}^m \lambda_{kj} f_j + u_k$ ,

$$\begin{aligned} \text{Cov}(x_i, x_k) &= \mathbb{E} \left[ \left( \sum_{j=1}^m \lambda_{ij} f_j + u_i \right) \left( \sum_{j=1}^m \lambda_{kj} f_j + u_k \right) \right] \\ &= \sum_{j=1}^m \sum_{l=1}^m \lambda_{ij} \lambda_{kl} \underbrace{\mathbb{E}(f_j f_l)}_{\delta_{jl}} + \sum_{j=1}^m \lambda_{ij} \underbrace{\mathbb{E}(f_j u_k)}_0 + \sum_{j=1}^m \lambda_{kj} \underbrace{\mathbb{E}(u_i f_j)}_0 + \underbrace{\mathbb{E}(u_i u_k)}_0 \\ &= \sum_{j=1}^m \lambda_{ij} \lambda_{kj}, \end{aligned}$$

donde hemos usado H2 ( $\mathbb{E}(f_j f_l) = \delta_{jl}$ ), H5 y H4 ( $\mathbb{E}(u_i u_k) = 0$  para  $i \neq k$ ).  $\square$

**Proposición 3.2** (Covarianza entre variables y factores).

$$\text{Cov}(\vec{x}, \vec{f}) = \Lambda.$$

Es decir, el peso factorial  $\lambda_{ij}$  es la covarianza entre la variable  $x_i$  y el factor  $f_j$ .

*Demostración.*

$$\text{Cov}(\vec{x}, \vec{f}) = \mathbb{E}[(\vec{x} - \vec{\mu}) \vec{f}^t] = \mathbb{E}[(\Lambda \vec{f} + \vec{u}) \vec{f}^t] = \Lambda \underbrace{\mathbb{E}(\vec{f} \vec{f}^t)}_{I_m} + \underbrace{\mathbb{E}(\vec{u} \vec{f}^t)}_0 = \Lambda. \quad \square$$

## 4. Unicidad del Modelo Factorial

### 4.1. Fuentes de indeterminación

El modelo factorial no es único. Existen dos tipos de indeterminación:

1. **Indeterminación por correlación entre factores:** ya hemos visto (Proposición 2) que los factores pueden estar correlados o incorrelados sin cambiar la matriz  $\Sigma$ .
2. **Indeterminación por rotación:** si  $C$  es cualquier matriz ortogonal  $m \times m$ , el modelo con  $\Lambda^* = \Lambda C^t$  y  $\vec{f}^* = C \vec{f}$  es indistinguible del original.

**Teorema 4.1** (Indeterminación por rotación). *Sea  $C \in \mathbb{R}^{m \times m}$  cualquier matriz ortogonal ( $CC^t = I_m$ ). Define  $\Lambda^* = \Lambda C^t$  y  $\vec{f}^* = C \vec{f}$ . Entonces el modelo  $\vec{x} = \vec{\mu} + \Lambda^* \vec{f}^* + \vec{u}$  satisface todas las hipótesis **H1–H5** y produce la misma descomposición de  $\Sigma$ .*

*Demostración.* Verificamos que el nuevo modelo satisface H2:

$$\text{Cov}(\vec{f}^*) = C \text{Cov}(\vec{f}) C^t \stackrel{\text{H2}}{=} C I_m C^t = CC^t = I_m.$$

Y que la descomposición de  $\Sigma$  se preserva:

$$\Lambda^* (\Lambda^*)^t + \Psi = \Lambda C^t C \Lambda^t + \Psi = \Lambda (C^t C) \Lambda^t + \Psi = \Lambda \Lambda^t + \Psi = \Sigma. \quad \square$$

## 4.2. Condiciones de unicidad: el criterio $\Lambda^t \Lambda$ diagonal

Para eliminar la indeterminación por rotación, se imponen restricciones adicionales sobre  $\Lambda$ .

**Definición 4.1** (Criterio de unicidad). Se dice que  $\Lambda$  satisface el **criterio canónico** si  $\Lambda^t \Lambda$  es diagonal con entradas decrecientes.

**Teorema 4.2** (Existencia y unicidad bajo el criterio canónico). *Si  $\Sigma - \Psi = \Lambda \Lambda^t$  y  $\Lambda^t \Lambda$  no es diagonal, existe una rotación  $C$  ortogonal tal que  $\Lambda^* = \Lambda C^t$  satisface  $(\Lambda^*)^t \Lambda^* = D$  diagonal. Además, ninguna otra rotación ortogonal no trivial de  $\Lambda^*$  produce una matriz diagonal.*

*Demostración.* Como  $\Lambda^t \Lambda$  es simétrica y semidefinida positiva, admite la descomposición espectral

$$\Lambda^t \Lambda = U D U^t,$$

con  $U$  ortogonal y  $D$  diagonal. Definimos  $C = U^t$  (ortogonal). Entonces

$$(\Lambda^*)^t \Lambda^* = (C^t)^t \Lambda^t \Lambda C^t = U \Lambda^t \Lambda U^t = U U D U^t U^t = D,$$

que es diagonal. Para la unicidad, supongamos que  $\Lambda^{**} = \Lambda^* C'$  para alguna matriz ortogonal  $C'$ . Entonces

$$(\Lambda^{**})^t \Lambda^{**} = (C')^t (\Lambda^*)^t \Lambda^* C' = (C')^t D C'.$$

Esta expresión es diagonal si y solo si  $C'$  diagonaliza  $D$ , lo que, siendo  $D$  ya diagonal, ocurre solo cuando  $C'$  es una matriz de permutaciones o de signos (matrices de permutación con entradas  $\pm 1$ ), que corresponden a reordenar y cambiar el signo de los factores, ambigüedad irreducible y sin consecuencias interpretativas.  $\square$

**Observación 4.1.** El criterio alternativo  $\Lambda^t \Psi^{-1} \Lambda$  diagonal conduce al método de máxima verosimilitud y es el preferido en estimación bajo normalidad.

## 5. Número Máximo de Factores

**Teorema 5.1** (Cota del número de factores). *El número máximo de factores  $m$  admisible en el modelo factorial con  $p$  variables es el mayor entero  $m$  tal que*

$$p(m+1) - \frac{m(m-1)}{2} \leq \frac{p(p+1)}{2}.$$

*Demostración.* La matriz de covarianzas  $\Sigma$  tiene  $\frac{p(p+1)}{2}$  parámetros libres (es simétrica). Por otro lado, el modelo  $\Sigma = \Lambda \Lambda^t + \Psi$  tiene:

- $pm$  parámetros en  $\Lambda$  (matriz  $p \times m$ ),
- $p$  parámetros en  $\Psi$  (diagonal),

es decir,  $p(m+1)$  parámetros en total. Sin embargo, la condición  $\Lambda^t \Lambda$  diagonal impone que  $\frac{m(m-1)}{2}$  elementos fuera de la diagonal sean cero, reduciendo el número efectivo de parámetros a

$$p(m+1) - \frac{m(m-1)}{2}.$$

Para que el sistema de ecuaciones  $\Sigma = \Lambda\Lambda^t + \Psi$  sea compatible determinado (o sobredeterminado), el número de ecuaciones debe ser al menos igual al de incógnitas:

$$p(m+1) - \frac{m(m-1)}{2} \leq \frac{p(p+1)}{2}. \quad \square$$

**Corolario 5.1.** *La diferencia  $\frac{p(p+1)}{2} - \left[ p(m+1) - \frac{m(m-1)}{2} \right] = \frac{(p-m)^2 - (p+m)}{2}$  es positiva (sistema sobredeterminado) siempre que  $p > m$ , lo que es precisamente la hipótesis del modelo.*

*Demostración.*

$$\begin{aligned} \frac{p(p+1)}{2} - p(m+1) + \frac{m(m-1)}{2} &= \frac{p^2 + p - 2pm - 2p + m^2 - m}{2} \\ &= \frac{p^2 - 2pm + m^2 - p - m}{2} \\ &= \frac{(p-m)^2 - (p+m)}{2}. \end{aligned}$$

Esta cantidad es positiva cuando  $(p-m)^2 > p+m$ , lo que se verifica para  $p$  suficientemente mayor que  $m$ . □

## 6. Estimación: El Método del Factor Principal

### 6.1. Planteamiento

A partir de una muestra de  $n$  observaciones del vector  $\vec{x}$ , se estima  $\Sigma$  mediante la matriz de varianzas-covarianzas muestral:

$$S = \frac{1}{n} X^t H X,$$

donde  $X$  es la matriz de datos (centrada) y  $H = I_n - \frac{1}{n} \mathbf{1}\mathbf{1}^t$  es la matriz de centrado.

El objetivo es estimar  $\Lambda$  (y  $\Psi$ ) a partir de  $S - \hat{\Psi} = \Lambda\Lambda^t$ .

### 6.2. Algoritmo iterativo

**Teorema 6.1** (Estimador del factor principal). *La solución al sistema  $S - \hat{\Psi} = \Lambda\Lambda^t$  bajo el criterio  $\Lambda^t\Lambda$  diagonal es*

$$\hat{\Lambda} = U_1 D_1^{1/2},$$

donde  $U_1$  es la matriz  $p \times m$  formada por los  $m$  vectores propios asociados a los valores propios no nulos de  $S - \hat{\Psi}$ , y  $D_1^{1/2}$  es la raíz cuadrada de la matriz diagonal de dichos valores propios.

*Demostración.* Como  $S - \hat{\Psi} = \Lambda\Lambda^t$  es simétrica y de rango  $m$  (pues  $\Lambda$  tiene dimensión  $p \times m$  con  $m < p$ ), admite la descomposición espectral

$$S - \hat{\Psi} = U D U^t = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_1^t \\ U_2^t \end{pmatrix} = U_1 D_1 U_1^t,$$

donde  $D_1$  contiene los  $m$  valores propios no nulos y  $U_1$  sus correspondientes vectores propios (ortonormales). Escribiendo  $D_1 = D_1^{1/2} D_1^{1/2}$ :

$$S - \hat{\Psi} = U_1 D_1^{1/2} D_1^{1/2} U_1^t = \hat{\Lambda} \hat{\Lambda}^t,$$

con  $\hat{\Lambda} = U_1 D_1^{1/2}$ .

Verificamos el criterio de unicidad:

$$\hat{\Lambda}^t \hat{\Lambda} = D_1^{1/2} U_1^t U_1 D_1^{1/2} = D_1^{1/2} I_m D_1^{1/2} = D_1,$$

que es diagonal. □

### 6.3. Algoritmo completo

El proceso de estimación es iterativo porque  $\Psi$  es desconocida:

#### 1. Inicialización de $\hat{\Psi}$ :

- *Opción A:*  $\hat{\psi}_{ii} = 0$  para todo  $i$  (equivalente a las componentes principales de  $S$ ; puede introducir mucho sesgo).
- *Opción B:*  $\hat{\psi}_{ii} = S_{ii}(1 - R_i^2)$ , donde  $R_i^2$  es el coeficiente de determinación de  $x_i$  en la regresión sobre todas las demás variables. Equivalentemente,  $\hat{\psi}_{ii} = \frac{1}{(S^{-1})_{ii}}$ . Justificación: Los factores comunes son los únicos que intervienen en todas las variables  $x_i$ ; es razonable estimar la comunalidad  $h_i^2 = S_{ii} - \psi_{ii}$  mediante la variación de  $x_i$  que puede explicarse con el resto:  $\hat{h}_i^2 = S_{ii} R_i^2$ .

#### 2. Primera estimación de $\hat{\Lambda}$ : aplicar el Teorema anterior a $S - \hat{\Psi}$ , tomando los $m$ mayores valores propios.

#### 3. Actualización de $\hat{\Psi}$ :

$$\hat{\Psi} = S - \hat{\Lambda} \hat{\Lambda}^t.$$

#### 4. Convergencia: repetir los pasos 2–3 hasta que la diferencia entre iteraciones consecutivas de $\hat{\Psi}$ sea inferior a un umbral prefijado.

## 7. Rotación de Factores

### 7.1. Motivación

Dado que el modelo factorial es indeterminado frente a rotaciones ortogonales, la solución obtenida por el algoritmo anterior puede no tener una interpretación clara. La idea es buscar una rotación  $C$  tal que la nueva matriz  $\Lambda^* = \hat{\Lambda} C^t$  tenga cargas factoriales con valores muy altos para ciertas variables y muy bajos para otras, facilitando la identificación de los factores con grupos de variables.

## 7.2. Criterio varimax

**Definición 7.1** (Criterio varimax). El **criterio varimax** (Kaiser, 1958) busca la matriz ortogonal  $C$  que maximiza la varianza de los cuadrados de las cargas factoriales normalizadas por columnas. Formalmente, se maximiza

$$V = \sum_{j=1}^m \text{Var}(\lambda_{1j}^{*2}, \dots, \lambda_{pj}^{*2}) = \sum_{j=1}^m \left[ \frac{1}{p} \sum_{i=1}^p \lambda_{ij}^{*4} - \left( \frac{1}{p} \sum_{i=1}^p \lambda_{ij}^{*2} \right)^2 \right],$$

donde  $\lambda_{ij}^* = (\Lambda^*)_{ij}$  son las cargas de  $\Lambda^* = \hat{\Lambda}C^t$ .

**Observación 7.1.** Maximizar la varianza de los cuadrados de las cargas equivale a concentrar los valores altos y bajos: se obtiene una solución donde cada factor tiene cargas próximas a  $\pm 1$  para un subconjunto de variables y próximas a 0 para las restantes, lo que facilita su interpretación temática.

**Observación 7.2.** El criterio varimax preserva las communalidades, pues la rotación es ortogonal:

$$\sum_{j=1}^m (\lambda_{ij}^*)^2 = \sum_{j=1}^m \lambda_{ij}^2 = h_i^2.$$

## 8. Resumen del Procedimiento

El procedimiento completo del análisis factorial puede resumirse en los siguientes pasos:

1. **Decidir el número de factores  $m$ :** por la cota del Teorema 5, o mediante el criterio de Kaiser (retener los factores con valor propio  $> 1$ ), o por el gráfico de sedimentación (*scree plot*).
2. **Estimar la matriz de cargas  $\hat{\Lambda}$ :** mediante el método del factor principal (o máxima verosimilitud).
3. **Verificar la bondad del ajuste:** comprobar que  $S \approx \hat{\Lambda}\hat{\Lambda}^t + \hat{\Psi}$  y que las communalidades son razonables.
4. **Rotar la solución:** aplicar el criterio varimax (u otro criterio de rotación) para mejorar la interpretabilidad.
5. **Interpretar los factores:** asignar un nombre o significado a cada factor en función de las variables con mayor carga en él.

## A. Apéndice: Propiedades de la Raíz Cuadrada de una Matriz

**Lema A.1** (Raíz cuadrada simétrica). Sea  $A \in \mathbb{R}^{n \times n}$  simétrica definida positiva con descomposición espectral  $A = UDU^t$ . Entonces la matriz  $A^{1/2} = UD^{1/2}U^t$  (con  $D^{1/2}$  la raíz cuadrada diagonal) satisface:

(a)  $A^{1/2}A^{1/2} = A$ .

(b)  $A^{1/2}$  es simétrica.

(c)  $A^{-1/2} = UD^{-1/2}U^t$  satisface  $A^{-1/2}A^{1/2} = I_n$ .

*Demostración.* (a)  $A^{1/2}A^{1/2} = UD^{1/2}U^t \cdot UD^{1/2}U^t = UD^{1/2}I_nD^{1/2}U^t = UDU^t = A$ .

(b)  $(A^{1/2})^t = (UD^{1/2}U^t)^t = (U^t)^t(D^{1/2})^tU^t = UD^{1/2}U^t = A^{1/2}$ , usando que  $U^t$  es ortogonal y  $D^{1/2}$  es diagonal (por tanto simétrica).

(c)  $A^{-1/2}A^{1/2} = UD^{-1/2}U^t \cdot UD^{1/2}U^t = UD^{-1/2}D^{1/2}U^t = UI_nU^t = I_n$ . □