

# ANALISIS FACTORIAL

## 0) MOTIVACION DEL MODELO

### Primera etapa- un solo factor (Charles Spearman)

A principios del siglo pasado los psicólogos estaban interesados en medir la inteligencia. Si bien este concepto es difícil de definir, se perciben claramente sus consecuencias. Los "inteligentes" resuelven mejor y más rápido problemas muy diversos, que pueden ser de tipo matemático, verbal, artístico, sociales, etc.

Charles Spearman se interesó por este tema aproximadamente por el año 1904.

El razonó que si los individuos tienen diferentes cantidades de inteligencia, y esta se "refleja" en la forma en que resuelven diversos problemas, una forma de medir la inteligencia sería a través de los puntajes que obtengan en ciertas pruebas.

El postuló entonces que si se administra un test (o prueba) a un individuo, el puntaje o score que obtenga,  $x$ , dependerá de su inteligencia  $I$ , según la relación lineal

$$x = \gamma I + \varepsilon$$

donde  $\varepsilon$  es un error aleatorio,  $\gamma$  es un coeficiente que depende del test, e  $I$  es la inteligencia del individuo.

Y si en lugar de un test, se administran  $d$  tests a cada individuo se tendrá

$$x_1 = \gamma_1 I + \varepsilon_1$$

$$x_2 = \gamma_2 I + \varepsilon_2$$

.....

$$x_d = \gamma_d I + \varepsilon_d$$

Notar que los  $\gamma_i$  son diferentes, ya que los tests tendrán en general distintas dificultades y escala, y también los  $\varepsilon_i$ .

Lo que sí es importante es que la inteligencia  $I$  es la misma, ya que depende del individuo, y es el mismo individuo el que resuelve todos los tests.

Dijimos poco respecto de  $I$ , pero ya que depende del individuo es razonable suponer que es una variable aleatoria, y podría postularse que

$$E(I) = \mu \quad \text{var}(I) = \sigma^2$$

y también que  $E(\varepsilon_i) = 0 \quad \text{var}(\varepsilon_i) = \psi_i^2$

También es lógico suponer que  $I$  y  $\varepsilon_i$ , estén no correlacionadas, ya que queremos que toda la inteligencia para resolver un test, esté en  $I$ , y no en  $\varepsilon_i$ .

Como último requisito pediremos que las  $\varepsilon_i$  estén no correlacionadas entre sí. El motivo es también sería deseable que las correlaciones entre los puntajes de dos tests,  $x_i$ , y  $x_j$ , dependan solamente de la inteligencia (a través de las  $\gamma_i$ ), y de nada más.

Este es el modelo de un factor de Spearman.

Escrito en forma matricial sería

$$\mathbf{x} = \Gamma \mathbf{I} + \boldsymbol{\varepsilon}$$

donde  $\mathbf{x}' = (x_1, x_2, \dots, x_d)$   $\Gamma' = (\gamma_1, \gamma_2, \dots, \gamma_d)$  y  $\boldsymbol{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d)$

Operativamente, para efectuar las estimaciones, hay que tomar una muestra de  $n$  individuos, hacerles resolver a cada uno los  $d$  tests, y con procedimientos que se verán más adelante se obtendrían estimaciones de :

- ♦ los  $\gamma_i$  que dependen de los tests
- ♦ los desvíos de los  $\varepsilon_i$ , o sea los  $\psi_i$  y
- ♦  $\mu$  y  $\sigma$  de la variable aleatoria inteligencia de la población a la que pertenecen los individuos.

Pero, y esto también se verá más adelante, se pueden estimar las inteligencias de los individuos.

Concretamente, para el individuo  $i$ , que resolvió los tests  $x_i$ , se puede estimar su inteligencia  $I_i$ .

Justamente esta era la intención de Spearman al abordar esta investigación.

## Segunda etapa - varios factores (L.L. Thurstone)

El modelo de Spearman para la inteligencia tropezó con dificultades.

La principal fué el reconocimiento por parte de los psicólogos, que la inteligencia no es un escalar, es decir no se la puede cuantificar por un solo número.

Tiene dimensiones. O sea debemos medirla con un vector.

Thurstone partió de considerar que un individuo, en lugar de tener cierta "cantidad" de inteligencia, tiene aptitudes: matemática, verbal, musical, etc., que ahora llamaremos factores.

El postuló entonces que si se administra un test (o prueba) a un individuo, el puntaje o score que obtenga,  $x$ , dependerá de las aptitudes del individuo **que se requieran para resolver ese test**. Por ejemplo para cierto test  $x_1$  podría ser que

$$x_1 = \gamma_{1m} \mathbf{M} + \gamma_{1v} \mathbf{V} + \gamma_{1s} \mathbf{S} + \varepsilon_1$$

o sea el puntaje dependería solo de las aptitudes matemáticas, verbales, y sociales, mas cierto error.

Sin embargo para otro test podría ocurrir que el modelo apropiado fuese

$$x_2 = \gamma_{2m} \mathbf{M} + \gamma_{2a} \mathbf{A} + \varepsilon_2$$

y esto querría decir que este test requiere solo de las aptitudes matemáticas y artísticas del individuo.

Si administramos  $d$  tests a cada individuo podríamos tener por ejemplo

$$\begin{aligned} x_1 &= \gamma_{1m} \mathbf{M} + \gamma_{1v} \mathbf{V} + \gamma_{1s} \mathbf{S} && + \varepsilon_1 \\ x_2 &= \gamma_{2m} \mathbf{M} && + \gamma_{2a} \mathbf{A} && + \varepsilon_2 \\ x_3 &= \gamma_{3m} \mathbf{M} + \gamma_{3v} \mathbf{V} && + \varepsilon_3 \\ &\dots\dots\dots && && \\ x_d &= && + \gamma_{dv} \mathbf{V} && + \gamma_{dr} \mathbf{R} && + \varepsilon_d \end{aligned}$$

Notar aquí que las aptitudes del individuo que se requieren para resolver esta batería de  $d$  tests, son de dos tipos:

- ♦ aptitudes "comunes", que son útiles para resolver 2 o más tests, como  $\mathbf{M}$  y  $\mathbf{V}$
- ♦ aptitudes "específicas", que intervienen solamente en un solo test, como  $\mathbf{S}$ ,  $\mathbf{A}$ , y  $\mathbf{R}$ .

Pero si en cada individuo hacemos solo una medición de cada test, no vamos a poder separar las componentes de aptitud específica del error de medición.

Luego, incorporaremos las aptitudes específicas de cada test, dentro del término de error.

Luego este modelo me permitirá explicar los puntajes de los tests en base a las aptitudes "comunes". Se tendrá entonces

$$\begin{aligned} x_1 &= \gamma_{1m} \mathbf{M} + \gamma_{1v} \mathbf{V} + \varepsilon_1 \\ x_2 &= \gamma_{2m} \mathbf{M} + \varepsilon_2 \end{aligned}$$

$$\begin{aligned}
 x_3 &= \gamma_{3m} \mathbf{M} + \gamma_{3v} \mathbf{V} + \varepsilon_3 \\
 &\dots\dots\dots \\
 x_d &= \dots\dots\dots + \gamma_{dv} \mathbf{V} + \varepsilon_d
 \end{aligned}$$

Donde los  $\varepsilon_i$  contienen tanto el error de medición como las aptitudes específicas del test. Hasta aquí hemos planteado un modelo en donde especificamos concretamente cuáles son las aptitudes "comunes" que intervienen en la resolución de la batería de tests. Sin embargo en la práctica esto suele desconocerse. Y este es justamente uno de los objetivos del análisis factorial, o sea descubrir cuáles son las aptitudes "comunes" que intervienen en la resolución de los tests. Por eso Thurstone a las aptitudes comunes les llamó factores, y planteó el modelo (asumiendo  $m$  factores comunes y agregando una media):

$$\begin{aligned}
 x_1 &= \gamma_{11} \mathbf{f}_1 + \gamma_{12} \mathbf{f}_2 + \gamma_{13} \mathbf{f}_3 + \dots\dots + \gamma_{1m} \mathbf{f}_m + \mu_1 + \varepsilon_1 \\
 x_2 &= \gamma_{21} \mathbf{f}_1 + \gamma_{22} \mathbf{f}_2 + \gamma_{23} \mathbf{f}_3 + \dots\dots + \gamma_{2m} \mathbf{f}_m + \mu_2 + \varepsilon_2 \\
 x_3 &= \gamma_{31} \mathbf{f}_1 + \gamma_{32} \mathbf{f}_2 + \gamma_{33} \mathbf{f}_3 + \dots\dots + \gamma_{3m} \mathbf{f}_m + \mu_3 + \varepsilon_3 \\
 &\dots\dots\dots \\
 x_d &= \gamma_{d1} \mathbf{f}_1 + \gamma_{d2} \mathbf{f}_2 + \gamma_{d3} \mathbf{f}_3 + \dots\dots + \gamma_{dm} \mathbf{f}_m + \mu_d + \varepsilon_d
 \end{aligned}$$

Las condiciones que le pediremos a este modelo son:

- 1) los  $\mathbf{f}_i$  son variables aleatorias de media 0, y matriz de covarianza  $\Phi$ . (usualmente se toma  $\Phi = \mathbf{I}_m$ , por conveniencia, ya que se desea que cada factor explique conceptos no correlacionados)
- 2) los  $\varepsilon_i$  son variables aleatorias de media 0, y matriz de covarianza  $\Psi$  **diagonal**. Y esta no correlación entre los  $\varepsilon_i$ , como veremos es muy importante en análisis factorial.
- 3) los factores no correlacionados con los  $\varepsilon_i$  (ya que no se quiere que cada  $\varepsilon_i$ , explique del test  $x_i$ , lo que explican los factores)

Si evaluamos la varianza de un test (por simplicidad supondremos  $\Phi = \mathbf{I}_m$ )

$$\text{var}(x_i) = \sigma_i^2 = \underbrace{\gamma_{i1}^2 + \gamma_{i2}^2 + \gamma_{i3}^2 + \dots\dots + \gamma_{im}^2}_{\mathbf{h}_i} + \psi_{ii} = \mathbf{h}_i + \psi_{ii}$$

donde  $\mathbf{h}_i$  se llama **comunalidad** del test  $i$ , y es la parte de su varianza que es debida a los factores comunes.

Y la covarianza entre dos tests

$$\text{cov}(x_i; x_j) = \sigma_{ij} = \gamma_{i1}\gamma_{j1} + \gamma_{i2}\gamma_{j2} + \gamma_{i3}\gamma_{j3} + \dots\dots + \gamma_{im}\gamma_{jm}$$

Notar aquí la importancia que  $\Psi$  sea **diagonal**, ya que así las covarianzas entre los tests, son explicadas solamente en base a los factores comunes.

**Observación:**

Si se calcula la correlación entre dos tests

$$\rho_{ij} = \sigma_{ij} / \sigma_i \sigma_j = (\gamma_{i1}\gamma_{j1} + \gamma_{i2}\gamma_{j2} + \gamma_{i3}\gamma_{j3} + \dots\dots + \gamma_{im}\gamma_{jm}) / \sqrt{(\mathbf{h}_i + \psi_{ii})(\mathbf{h}_j + \psi_{jj})}$$

Virgilio L. Foglia (2003)

vemos que depende de los coeficientes de los factores comunes ( $\gamma_{ij}$ ), de las comunalidades de los tests (que también dependen de los  $\gamma_{ij}$ ), pero también de  $\psi_{ii}$  y  $\psi_{jj}$ .

Esto quiere decir que las correlaciones entre los tests, no son explicadas totalmente por los factores comunes.

Sin embargo, si los puntajes de los tests se miden estandarizados,  $\sigma_i^2 = \sigma_j^2 = 1$ , y entonces sí es cierto que las correlaciones entre los tests son totalmente explicadas por los factores comunes.

## Pasos y objetivos del análisis factorial

Usualmente para realizar un análisis factorial se necesita una muestra aleatoria de  $n$  individuos,

donde cada uno debe resolver  $d$  tests o pruebas, obteniéndose los puntajes correspondientes.

De aquí se estima la matriz de covarianza de los tests (o a veces la matriz de correlación), y se

aplican procedimientos de estimación y análisis que veremos más adelante. Los objetivos son:

- 1) Determinar  $m$ , el número de factores comunes que explican los puntajes de los tests.
- 2) Estimar las cargas de los factores (los  $\gamma_{ij}$ ). Esto permite analizar que factores o aptitudes intervienen en la resolución de cada test.
- 3) Interpretar el significado de los factores. Y esto como herramienta exploratoria es muy importante, ya que permite "descubrir" quiénes son los factores, cuando uno se encuentra en las primeras etapas de una investigación.. En esta etapa es útil emplear algún

procedimiento

para rotar los factores, y facilitar la interpretación.

- 4) Estimar los factores. O sea, en el caso de los tests, estimar "cuanto", de cada aptitud tiene cada individuo.

- 5) Evaluar la calidad del ajuste.

## Ejemplo

Supongamos que se dispone de un listado de calificaciones de fin de curso, de 200 estudiantes,

en Matemática (M), Física (F), Química (Q), Inglés (I), Historia(H), Alemán (A).

Calculando la matriz de correlación muestral resulta

$$S = \begin{bmatrix} 1 & 0.62 & 0.54 & 0.32 & 0.284 & 0.37 \\ 0.62 & 1 & 0.51 & 0.38 & 0.351 & 0.43 \\ 0.54 & 0.51 & 1 & 0.36 & 0.336 & 0.405 \\ 0.32 & 0.38 & 0.36 & 1 & 0.686 & 0.73 \\ 0.284 & 0.351 & 0.336 & 0.686 & 1 & 0.735 \\ 0.37 & 0.43 & 0.405 & 0.73 & 0.735 & 1 \end{bmatrix}$$

Suponiendo  $m = 2$  factores, y aplicando el método de estimación por MV resultan estimados los coeficientes de los factores

$$M = 0.572 \mathbf{f}_1 + 0.594 \mathbf{f}_2 + \varepsilon_1$$

Virgilio L. Foglia (2003)

$$\begin{aligned}F &= 0.609 \mathbf{f}_1 + 0.458 \mathbf{f}_2 + \varepsilon_2 \\Q &= 0.559 \mathbf{f}_1 + 0.371 \mathbf{f}_2 + \varepsilon_3 \\I &= 0.793 \mathbf{f}_1 - 0.224 \mathbf{f}_2 + \varepsilon_4 \\H &= 0.786 \mathbf{f}_1 - 0.278 \mathbf{f}_2 + \varepsilon_5 \\A &= 0.862 \mathbf{f}_1 - 0.206 \mathbf{f}_2 + \varepsilon_6\end{aligned}$$

La interpretación aquí de los factores sería:

$\mathbf{f}_1 \rightarrow$  un factor general de inteligencia

$\mathbf{f}_2 \rightarrow$  un factor de aptitud que diferencia entre habilidad cuantitativa y verbal

Las comunales de cada variable serían:

$$\begin{aligned}h_M &= 0.572^2 + 0.594^2 = 0.68 \\h_F &= 0.609^2 + 0.458^2 = 0.58 \\h_Q &= 0.559^2 + 0.371^2 = 0.45 \\h_I &= 0.793^2 + (-0.224)^2 = 0.68 \\h_H &= 0.786^2 + (-0.278)^2 = 0.695 \\h_A &= 0.862^2 + (-0.206)^2 = 0.785\end{aligned}$$

Estas miden el aporte que hacen los factores comunes en la explicación de cada variable.  
La varianza total explicada por cada factor

$$\begin{aligned}\sigma_{\mathbf{f}_1}^2 &= 0.572^2 + 0.609^2 + 0.559^2 + 0.793^2 + 0.786^2 + 0.862^2 = 2.999 \\ \sigma_{\mathbf{f}_2}^2 &= 0.594^2 + 0.458^2 + 0.371^2 + (-0.224)^2 + (-0.278)^2 + (-0.206)^2 = 0.871\end{aligned}$$

Estas miden la importancia de cada factor en la explicación de todas las variables.  
Si como veremos más adelante, para mejorar la interpretación de los factores, se los rota (se empleó aquí el método varimax), quedará para el modelo rotado

$$\begin{aligned}M &= 0.151 \mathbf{f}_1^* + \mathbf{0.811} \mathbf{f}_2^* + \varepsilon_1 \\F &= 0.257 \mathbf{f}_1^* + \mathbf{0.717} \mathbf{f}_2^* + \varepsilon_2 \\Q &= 0.263 \mathbf{f}_1^* + \mathbf{0.617} \mathbf{f}_2^* + \varepsilon_3 \\I &= \mathbf{0.786} \mathbf{f}_1^* + 0.248 \mathbf{f}_2^* + \varepsilon_4 \\H &= \mathbf{0.810} \mathbf{f}_1^* + 0.199 \mathbf{f}_2^* + \varepsilon_5 \\A &= \mathbf{0.834} \mathbf{f}_1^* + 0.301 \mathbf{f}_2^* + \varepsilon_6\end{aligned}$$

Ahora sí la interpretación es más fácil

$\mathbf{f}_1^* \rightarrow$  un factor que mide aptitud verbal

$\mathbf{f}_2^* \rightarrow$  un factor que mide aptitud cuantitativa

Las comunales de las variables serán las mismas, o sea

$$\begin{aligned}h_M &= 0.151^2 + 0.811^2 = 0.68 \\h_F &= 0.257^2 + 0.717^2 = 0.58 \\h_Q &= 0.263^2 + 0.617^2 = 0.45 \\h_I &= 0.786^2 + 0.248^2 = 0.68 \\h_H &= 0.810^2 + 0.199^2 = 0.695 \\h_A &= 0.834^2 + 0.301^2 = 0.785\end{aligned}$$

Pero ahora la varianza total explicada por cada factor cambiará

$$\sigma_{f_1}^2 = 0.151^2 + 0.257^2 + 0.263^2 + 0.786^2 + 0.810^2 + 0.834^2 = 2.128$$

$$\sigma_{f_2}^2 = 0.811^2 + 0.717^2 + 0.617^2 + 0.248^2 + 0.199^2 + 0.301^2 = 1.743$$

## 1) PLANTEO DEL MODELO

### Modelo con factores aleatorios

$$\boxed{\mathbf{x} = \Gamma \mathbf{f} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}}$$
 donde  $\mathbf{x}$ ,  $\boldsymbol{\mu}$ , y  $\boldsymbol{\varepsilon}$  están en  $\mathbf{R}^d$ ,  $\mathbf{f}$  en  $\mathbf{R}^m$ , y  $\Gamma$  en  $\mathbf{R}^{d \times m}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_d \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} \quad \Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \gamma_{21} & \cdots & \gamma_{2m} \\ \vdots & & \vdots \\ \gamma_{d1} & \cdots & \gamma_{dm} \end{bmatrix}$$

$\mathbf{x}$  es un vector aleatorio **observable**;  $\mathbf{f}$ ,  $\boldsymbol{\mu}$ , y  $\boldsymbol{\varepsilon}$  son vectores aleatorios **no observables**

Se pide:

- 1)  $\mathbf{f}$  es un vector aleatorio con  $E(\mathbf{f}) = \mathbf{0}$      $\text{cov}(\mathbf{f}) = \Phi$
- 2)  $\boldsymbol{\varepsilon}$  es un vector aleatorio con  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$      $\text{cov}(\boldsymbol{\varepsilon}) = \Psi$  (**diagonal**)
- 3)  $\text{cov}(\mathbf{f}; \boldsymbol{\varepsilon}) = \mathbf{0}$

4)  $\text{rg}(\Gamma) = \mathbf{m}$

(esta última condición se pide ya que si  $\text{rg}(\Gamma) = \mathbf{r} < \mathbf{m}$ , siempre existirá  $\Gamma_r \in \mathbf{R}^{d \times r}$ , de rango  $r$ , con  $\Gamma_r \Gamma_r' = \Gamma \Gamma'$ , y entonces  $\mathbf{x} = \Gamma_r \mathbf{f}_r + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$  sería otra solución con  $\Phi = \mathbf{I}_r$  e igual estructura de covarianza).

Además si  $\mathbf{x} = \Gamma \mathbf{f} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$  resulta

$$\Sigma = \text{cov}(\mathbf{x}; \mathbf{x}) = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = E[(\Gamma \mathbf{f} + \boldsymbol{\varepsilon})(\Gamma \mathbf{f} + \boldsymbol{\varepsilon})'] = \underbrace{\Gamma E(\mathbf{f} \mathbf{f}') \Gamma'}_{\Phi} + \underbrace{E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}')}_{\Psi}$$

Luego queda

$$\boxed{\Sigma = \Gamma \Phi \Gamma' + \Psi}$$

#### Simplificación del modelo: (requisito 1)

$$\text{Como } \mathbf{x} = \Gamma \mathbf{f} + \boldsymbol{\mu} + \boldsymbol{\varepsilon} = \underbrace{\Gamma \Phi^{1/2}}_{\Gamma^*} \underbrace{\Phi^{-1/2} \mathbf{f}}_{\mathbf{f}^*} + \boldsymbol{\mu} + \boldsymbol{\varepsilon} = \Gamma^* \mathbf{f}^* + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

$$\text{con } \text{cov}(\mathbf{f}^*) = E(\mathbf{f}^* \mathbf{f}^{*'}) = E(\Phi^{-1/2} \mathbf{f} \mathbf{f}' \Phi^{-1/2}) = \Phi^{-1/2} \Phi \Phi^{-1/2} = \mathbf{I}_m$$

Luego un modelo con  $\Gamma$ ,  $\mathbf{f}$ , y  $\Phi$  lo podemos expresar como otro con  $\Gamma^*$ ,  $\mathbf{f}^*$ , y  $\mathbf{I}_m$ .

Entonces sin pérdida de generalidad supondremos en nuestro modelo que

$$\boxed{\Phi = \mathbf{I}_m} \quad \rightarrow \quad \boxed{\Sigma = \Gamma \Phi \Gamma' + \Psi}$$

(ya que, si después pensamos que es mejor un modelo con factores correlacionados, será cuestión de tomar  $\Gamma^* = \Gamma \Phi^{1/2}$  y  $\mathbf{f}^* = \Phi^{-1/2} \mathbf{f}$ , y listo).

#### Principio de independencia condicional (requisito 2)

La condición  $\text{cov}(\boldsymbol{\varepsilon}) = \Psi$  (**diagonal**), en el caso de normalidad, implica la validez del

principio de independencia condicional

”Si los factores son conocidos, las observaciones  $\mathbf{x}$  resultan independientes”

Luego, toda la información respecto de las relaciones entre las variables observadas se encuentra en los factores.

Este requisito  $\text{cov}(\varepsilon) = \Psi$  (**diagonal**) es fundamental en análisis factorial, ya que el objetivo esencial de este método es describir las interrelaciones (covarianzas) entre las variables observadas, en función de unas pocas variables no observadas, llamadas factores.

### Una inquietud (requisito 2 y 3)

Recién planteamos el modelo  $\mathbf{x} = \Gamma \mathbf{f} + \mu + \varepsilon$  (supongamos  $\text{rg}(\Sigma) = \mathbf{d}$ ).

Podríamos preguntarnos si no es posible que los factores puedan expresarse en función de las  $\mathbf{x}$  así:

$$\mathbf{f} = \mathbf{B}\mathbf{x} \quad \text{con } \mathbf{B} \in \mathbb{R}^{m \times d} \text{ y } \text{rg}(\mathbf{B}) = \mathbf{m}$$

y entonces el modelo sería del tipo

$$\mathbf{x} = \Gamma \mathbf{B}\mathbf{x} + \mu + \varepsilon$$

o sea

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mu + \varepsilon \quad \text{con } \mathbf{A} \in \mathbb{R}^{d \times d} \text{ con } \text{rg}(\mathbf{A}) = \mathbf{m}$$

$$(\mathbf{I}_d - \mathbf{A})\mathbf{x} = \mu + \varepsilon$$

Luego la matriz de covarianza de  $\varepsilon$  sería

$$\Psi = (\mathbf{I}_d - \mathbf{A})\Sigma(\mathbf{I}_d - \mathbf{A})'$$

pero si exijo que  $\Psi$  sea diagonal, entonces  $(\mathbf{I}_d - \mathbf{A})' = \mathbf{T}$ , los autovectores de  $\Sigma$ , y  $\text{rg}(\Psi) = \mathbf{d}$

y la

$$\text{cov}(\mathbf{f}; \varepsilon) = \mathbf{E}(\mathbf{B}\mathbf{x}\mathbf{x}'(\mathbf{I}_d - \mathbf{A})') = \mathbf{B}\Sigma\mathbf{T} = \mathbf{B}\mathbf{T}\Psi$$

pero si también exijo que esta matriz de covarianza sea nula

$$\mathbf{B}\mathbf{T}\Psi = \mathbf{0} \rightarrow \mathbf{B}\mathbf{T} = \mathbf{0} \rightarrow \mathbf{B} = \mathbf{0} \text{ (abs.)}$$

Luego no puedo lograr tal descomposición.

De aquí que en análisis factorial, los factores, al no poderse expresar en función de las variables  $\mathbf{x}$  observadas, cumpliendo los requisitos  $\Psi$  diagonal, y  $\text{cov}(\mathbf{f}; \varepsilon) = \mathbf{0}$ , tengan que ser ”otras” variables aleatorias. Por eso se requiere introducir el concepto de variables **no observadas**, o **latentes**.

### Matriz de covarianza entre observaciones y factores

A veces suele interesar la matriz de covarianza entre las observaciones y los factores.

$$\text{cov}(\mathbf{x}; \mathbf{f}) = \mathbf{E}[(\mathbf{x} - \mu)\mathbf{f}'] = \mathbf{E}[(\Gamma\mathbf{f} + \varepsilon)\mathbf{f}'] = \Gamma \underbrace{\mathbf{E}(\mathbf{f}\mathbf{f}')}_{\Phi} + \underbrace{\mathbf{E}(\varepsilon\mathbf{f}')}_{\mathbf{0}}$$

Luego queda

$$\Pi = \Gamma\Phi$$

y para el caso en que  $\Phi = \mathbf{I}_m$

$$\Pi = \Gamma$$

### Indeterminación

Sea  $A = \left\{ \mathbf{A} : \mathbf{A} \in \mathbb{R}^{m \times m} \text{ no singular} \right\}$

Virgilio L. Foglia (2003)

$$\mathbf{x} = \Gamma \mathbf{f} + \mu + \varepsilon = \mathbf{x} = \underbrace{\Gamma \mathbf{A} \mathbf{A}^{-1}}_{\substack{\Gamma^* \\ \mathbf{f}^*}} \mathbf{f} + \mu + \varepsilon$$

Luego observando  $\mathbf{x}$ , no podemos distinguir entre un modelo con  $\Gamma, \mathbf{f}$ , y  $\Phi$  y otro con  $\Gamma^*, \mathbf{f}^*$ , y  $\Phi^* = \mathbf{A}^{-1} \Phi \mathbf{A}^{-1}$

Pero si supongo  $\Phi = \Phi^* = \mathbf{I}_m \rightarrow \mathbf{A}^{-1} \mathbf{A}^{-1'} = \mathbf{I}_m \rightarrow \mathbf{A} \mathbf{A}' = \mathbf{I}_m$   
Luego la indeterminación queda reducida a

$$A^+ = \left\{ \mathbf{A} : \mathbf{A} \in \mathbf{R}^{m \times m} \quad \mathbf{A} \mathbf{A}' = \mathbf{A}' \mathbf{A} = \mathbf{I}_m \right\} \text{ (todas las rotaciones de los factores)}$$

## Modelo con factores fijos

$$\boxed{\mathbf{x}_a = \Gamma \mathbf{f}_a + \mu + \varepsilon_a} \quad a = 1 \dots n \quad \text{donde } \mathbf{x}_a, \mu, \text{ y } \varepsilon_a \text{ están en } \mathbf{R}^d, \mathbf{f}_a \text{ en } \mathbf{R}^m, \text{ y } \Gamma \text{ en } \mathbf{R}^{d \times m}$$

Donde los  $\mathbf{f}_a$  son vectores **no aleatorios**, que son **parámetros**.

Podemos también pensar a las  $\mathbf{x}_a$  como variables condicionales, o sea  $\mathbf{x}_a \sim \mathbf{x} / \mathbf{f}$ .

Se pide:

- 1)  $\mathbf{f}_a$  es un vector **no aleatorio** de parámetros con  $\frac{1}{n} \sum \mathbf{f}_a = \mathbf{0} \quad \frac{1}{n} \sum \mathbf{f}_a \mathbf{f}_a' = \Theta$
- 2)  $\varepsilon_a$  es un vector aleatorio con  $E(\varepsilon_a) = \mathbf{0} \quad \text{cov}(\varepsilon_a) = \Psi$  (**diagonal**)

## Una interpretación

Supondremos para simplificar que  $\mu = \mathbf{0}$ , o sea consideramos las  $\mathbf{x}_a$  centradas.

Las matrices, de datos observados de las  $\mathbf{x}_a$ , y de factores parámetros  $\mathbf{f}_a$ ,

y de errores  $\varepsilon_a$ , para  $a = 1 \dots n$ , son

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & x_{31} & x_{41} & x_{51} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & x_{42} & x_{52} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ x_{1d} & x_{2d} & x_{3d} & x_{4d} & x_{5d} & \cdots & x_{nd} \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} & f_{51} & \cdots & f_{n1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ f_{1m} & f_{2m} & f_{3m} & f_{4m} & f_{5m} & \cdots & f_{nm} \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} & \varepsilon_{31} & \varepsilon_{41} & \varepsilon_{51} & \cdots & \varepsilon_{n1} \\ \varepsilon_{12} & \varepsilon_{22} & \varepsilon_{32} & \varepsilon_{42} & \varepsilon_{52} & \cdots & \varepsilon_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \varepsilon_{1d} & \varepsilon_{2d} & \varepsilon_{3d} & \varepsilon_{4d} & \varepsilon_{5d} & \cdots & \varepsilon_{nd} \end{bmatrix}$$

Luego queda:  $\mathbf{X} = \Gamma \mathbf{F} + \varepsilon \rightarrow \mathbf{X}' = \mathbf{F}' \Gamma' + \varepsilon'$



Virgilio L. Foglia (2003)

Si llamamos  $\mathbf{X}' = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_d)$  donde  $\mathbf{X}_i$  tiene las  $\mathbf{n}$  observaciones del test  $\mathbf{i}$

$\mathbf{F}' = (\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_m)$  donde  $\mathbf{F}_j$  tiene los  $\mathbf{n}$  parámetros del factor  $\mathbf{j}$

$\boldsymbol{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d)$  donde  $\varepsilon_i$  tiene las  $\mathbf{n}$  errores del test  $\mathbf{i}$

$\boldsymbol{\Gamma}' = (\Gamma_1, \Gamma_2, \dots, \Gamma_d)$  donde  $\Gamma_i$  tiene la fila  $\mathbf{i}$  de la matriz  $\boldsymbol{\Gamma}$

Luego  $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_d) = \mathbf{F}'(\Gamma_1, \Gamma_2, \dots, \Gamma_d) + (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d)$

Y resultan los modelos:

$$\mathbf{X}_1 = \mathbf{F}'\Gamma_1 + \varepsilon_1$$

$$\mathbf{X}_2 = \mathbf{F}'\Gamma_2 + \varepsilon_2$$

$$\mathbf{X}_3 = \mathbf{F}'\Gamma_3 + \varepsilon_3$$

.....

$$\mathbf{X}_d = \mathbf{F}'\Gamma_d + \varepsilon_d$$

Si los parámetros de  $\mathbf{F}'$  fuesen conocidos, claramente esto son  $\mathbf{d}$  modelos lineales.

## 2) EXISTENCIA (1er problema de la población)

El problema es cuándo una  $\Sigma \in \mathbf{R}^{d \times d}$  definida positiva, se puede expresar como

$$\Sigma = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi} \quad \text{con } \boldsymbol{\Psi} \text{ diagonal y } \boldsymbol{\Gamma} \in \mathbf{R}^{d \times m}$$

Si tengo  $\Sigma$ , para llegar a una solución tendría que resolver las siguientes  $\frac{d(d+1)}{2}$  ecuaciones en  $\mathbf{dm} + \mathbf{d}$  incógnitas

$$\sigma_{ii} = \sum_{k=1}^m \gamma_{ik}^2 + \Psi_{ii} \quad (\text{para } i = 1 \dots \mathbf{d}) \quad \sigma_{ij} = \sum_{k=1}^m \gamma_{ik}\gamma_{jk} \quad (\text{para } i < j)$$

Pero dada una solución  $\boldsymbol{\Gamma}$ , puede ser reemplazada por  $\boldsymbol{\Gamma}\mathbf{L}$  con  $\mathbf{L}$  ortogonal y  $\frac{m(m-1)}{2}$  elementos independientes.

O sea esto equivale a que  $\boldsymbol{\Gamma}$  satisfaga  $\frac{m(m-1)}{2}$  condiciones adicionales.

Luego si  $\mathbf{C} = \text{nro ecuaciones} + \text{nro condiciones} - \text{nro incógnitas}$

$$\mathbf{C} = \frac{d(d+1)}{2} + \frac{m(m-1)}{2} - \mathbf{dm} - \mathbf{d}$$

$$\mathbf{C} = \frac{(d-m)^2 - (d+m)}{2}$$

Si  $\mathbf{C} \leq 0$  uno esperaría que una solución algebraica sea posible.

Será cierto que  $\mathbf{C} \leq 0 \Rightarrow$  Existencia ? No, ya que la solución podría no ser real, o obtenerse algún  $\Psi_{ii} < 0$ .

Será cierto que  $\mathbf{C} \leq 0 \Leftarrow$  Existencia ? No, ya que no tenemos seguridad que las

ecuaciones sean independientes.

Resolviendo la inequación  $C \leq 0$  resulta que una condición poco rigurosa es que

$$\mathbf{m} \geq \mathbf{d} - \frac{\sqrt{8\mathbf{d}+1} - 1}{2}$$

O sea para  $d = 1 \quad m \geq 0$

$d = 3 \quad m \geq 1$

$d = 5 \quad m \geq 2.3$

$d = 6 \quad m \geq 3$

$d = 3 \quad m \geq 1$

$d = 10 \quad m \geq 6$

$d = 15 \quad m \geq 10$

■ Teorema :  $\Sigma$  admite  $\mathbf{m}$  factores  $\Leftrightarrow \exists \Psi \geq 0$  **diagonal**, tal que  $\Sigma - \Psi \geq 0$  de rango  $\mathbf{m}$

(ya que dada  $A \in \mathbf{R}^{d \times d}$ ,  $\exists B \in \mathbf{R}^{d \times m}$ ,  $\text{rg}(B) = \mathbf{m}$  con  $A = BB'$   $\Leftrightarrow A \geq 0$  de rango  $\mathbf{m}$ )

## Caso $m=1$

Aquí  $\Sigma$  debería satisfacer  $C = \frac{d(d+1)}{2} - 2d$  condiciones de igualdad y alguna desigualdad.

Según el teorema, debe existir  $\Psi \geq 0$ , de manera que  $\text{rg}(\Sigma - \Psi) = 1$  y  $1/2$  def.positiva.

Esto implica que en  $\Sigma - \Psi$  **todos los menores de 2do orden sean nulos**.

Se define en  $\Sigma$  : **tétrada**  $_{i j h k} = \begin{vmatrix} \sigma_{hi} & \sigma_{hj} \\ \sigma_{ki} & \sigma_{kj} \end{vmatrix}$  (con  $i j h k$  diferentes)

Obviamente todas las tétradas de  $\Sigma$  deben ser nulas.

Pero en  $\Sigma - \Psi$  tomemos un menor que incluya un elemento diagonal, deberemos exigirle

$$\begin{vmatrix} \sigma_{ii} - \Psi_{ii} & \sigma_{ij} \\ \sigma_{ki} & \sigma_{kj} \end{vmatrix} = 0 \Rightarrow \Psi_{ii} = \sigma_{ii} - \frac{\sigma_{ij}\sigma_{ki}}{\sigma_{kj}} \quad \text{y así hallo los } \Psi_{ii}$$

Se puede demostrar que si **todas** las tétradas de  $\Sigma$  son nulas:

1) Los  $\Psi_{ii}$  son consistentes (independientes de  $j k$ )

2) Los menores de  $\Sigma - \Psi$  con dos elementos en la diagonal también son nulos.

También se puede demostrar :

$$C = \frac{d(d+1)}{2} - 2d \text{ tétradas nulas} \Leftrightarrow \text{Todas las tétradas son nulas}$$

■ Teorema:  $\Sigma$  admite **1** factor  $\Leftrightarrow$  hay **C** tétradas nulas y  $0 \leq \frac{\sigma_{ij}\sigma_{ki}}{\sigma_{kj}} \leq \sigma_{ii}$  ( $j \neq k$ )  $\forall i$

Pues  $\Psi_{ii} = \sigma_{ii} - \frac{\sigma_{ij}\sigma_{ki}}{\sigma_{kj}} \geq 0$  pues  $\Psi \geq \mathbf{0}$

$$\frac{\sigma_{ij}\sigma_{ki}}{\sigma_{kj}} \geq 0 \text{ pues } \Sigma - \Psi \geq \mathbf{0}$$

Luego hemos encontrado una condición necesaria y suficiente a exigirle a  $\Sigma$  para que admita un modelo factorial con **1** factor.

■ Teorema de Reiersol (1950)

$\exists$  solución con **m** factores  $\Rightarrow \exists$  infinitas soluciones con **m+1** factores

Si  $\Sigma = \Gamma_m \Gamma_m' + \Psi_m = \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \Psi_m$  tomando  $0 < \theta < 1$  expreso

$$\Psi_m = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta\psi_{m+1m+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{e}_{m+1}\theta\psi_{m+1m+1}\mathbf{e}_{m+1}'} + \underbrace{\begin{bmatrix} \psi_{11} & 0 & 0 & 0 & 0 \\ 0 & \cdot & 0 & 0 & 0 \\ 0 & 0 & (1-\theta)\psi_{m+1m+1} & 0 & 0 \\ 0 & 0 & 0 & \cdot & 0 \\ 0 & 0 & 0 & 0 & \psi_{dd} \end{bmatrix}}_{\Psi_{m+1}}$$

Donde  $\mathbf{e}_{m+1}$  es un vector columna con ceros salvo un 1 en posición **m+1**, luego

$$\Sigma = \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \mathbf{e}_{m+1}\theta\psi_{m+1m+1}\mathbf{e}_{m+1}' + \Psi_{m+1}$$

$$\Sigma = \begin{bmatrix} \mathbf{T}_m & \mathbf{e}_{m+1} \end{bmatrix} \begin{bmatrix} \Lambda_m & 0 \\ 0 & \theta\psi_{m+1m+1} \end{bmatrix} \begin{bmatrix} \mathbf{T}_m & \mathbf{e}_{m+1} \end{bmatrix}' + \Psi_{m+1}$$

pero como  $\mathbf{e}_{m+1}$  es linealmente independiente de las columnas de  $\mathbf{T}_m$

$\text{rg}\left(\begin{bmatrix} \mathbf{T}_m & \mathbf{e}_{m+1} \end{bmatrix}\right) = \mathbf{m} + 1$ , luego existirá  $\Gamma_{m+1}$  de rango **m+1**, con

$$\Sigma = \Gamma_{m+1} \Gamma_{m+1}' + \Psi_{m+1}$$

y como  $\theta$  es arbitrario, existirán infinitas soluciones.

■ Corolario

$\exists$  solución con **m** factores **UNICA**  $\Rightarrow$  **m** es el **mínimo** posible

ya que si existe solución para un **m** menor, por el teorema, para **m** factores la solución no sería única.

### 3) IDENTIFICACION (2do problema de la población)

Vimos que si  $\mathbf{C} \leq \mathbf{0}$  uno esperaría que una solución algebraica sea posible (**EXISTENCIA**).

Luego si  $C \geq 0$  uno esperaría **IDENTIFICACION**.

En lo que sigue supondremos que **existe alguna** solución  $(\Gamma; \Psi)$  para  $\Sigma = \Gamma\Gamma' + \Psi$

(quizás sujeta a ciertas restricciones) y nos proponemos investigar **si hay otras**.

Podría ocurrir que existan otras soluciones:

- (a) con el mismo  $\Psi$ , como  $(\Gamma^*; \Psi) \Rightarrow$  Problema de identificación de **Rotación**
- (b) con el mismo  $\Gamma$ , como  $(\Gamma; \Psi^*) \Rightarrow$  (esto no puede ocurrir)
- (c) con diferente  $\Psi$  y  $\Gamma$ , como  $(\Gamma^*; \Psi^*) \Rightarrow$  Problema de identificación de **Estructura**

**Observación:** Cuando estudiamos el problema de existencia, buscábamos condiciones a exigir a  $\Sigma$ , para que exista una solución. Aquí en identificación, ya tenemos una solución  $(\Gamma, \Psi)$ , y por comodidad analizaremos que condiciones debe cumplir  $\Gamma$  para que la solución sea única.

## Identificación respecto de la Rotación.

Teorema Si  $(\Gamma; \Psi)$  es solución de rango  $m$

$$(\Gamma^*; \Psi) \text{ es solución} \Leftrightarrow \exists \mathbf{L} \text{ ortogonal con } \Gamma^* = \Gamma \mathbf{L}$$

$$\Leftrightarrow \Gamma^* \Gamma^{*'} + \Psi = \Gamma \mathbf{L} \mathbf{L}' \Gamma' + \Psi = \Gamma \Gamma' + \Psi = \Sigma \text{ luego es solución}$$

$$\Rightarrow \Gamma \Gamma' + \Psi = \Gamma^* \Gamma^{*'} + \Psi \rightarrow \Gamma \Gamma' = \Gamma^* \Gamma^{*'}$$

$$\text{pero } \mathbf{rg}(\Gamma) = m \rightarrow \mathbf{rg}(\Gamma \Gamma') = m \rightarrow \mathbf{rg}(\Gamma^*) = m, \text{ multiplicando por } \Gamma^* (\Gamma^* \Gamma^*)^{-1}$$

$$\Gamma^* \underbrace{\Gamma^{*' \Gamma^* (\Gamma^{*' \Gamma^*)^{-1}}}_{\mathbf{I}_m}} = \underbrace{\Gamma \Gamma' \Gamma^* (\Gamma^* \Gamma^*)^{-1}}_{\mathbf{L}}$$

$$\text{y } \mathbf{L}' \mathbf{L} = (\Gamma^{*' \Gamma^*})^{-1} \Gamma^{*' \underbrace{\Gamma \Gamma' \Gamma^* (\Gamma^* \Gamma^*)^{-1}}_{\Gamma^* \Gamma^{*'}}} = \mathbf{I}_m \rightarrow \text{ortogonal} \rightarrow \Gamma^* = \Gamma \mathbf{L}$$

De este teorema resulta que si hay una solución  $(\Gamma; \Psi)$ , cualquier otra de igual  $\Psi$  difiere en una rotación.

## Eliminación de la indeterminación de rotación

Resolvemos este problema de identificación, imponiéndole alguna restricción a  $\Gamma$ , (la única indeterminación que no eliminaremos es la que se refiere a multiplicar por -1 algunas columnas de  $\Gamma$ , ya que esto equivale a cambiarle el signo a los correspondientes factores).

Como en la indeterminación de rotación,  $\Psi$  se mantiene fija, debemos imponer alguna restricción a  $\Gamma$ , de manera que  $\Gamma \Gamma' = C$  se mantenga igual.

Las restricciones habituales son:

- (a) Forma triangular de  $\Gamma$

$$\Gamma = \begin{bmatrix} \gamma_{11} & 0 & 0 & \cdots & 0 \\ \gamma_{21} & \gamma_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & & \cdots & \\ \gamma_{m1} & \gamma_{m2} & \gamma_{m3} & \cdots & \gamma_{mm} \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{d1} & \gamma_{d2} & \gamma_{d3} & \cdots & \gamma_{dm} \end{bmatrix}$$

(b) Condición general de triangularidad

Si  $\mathbf{B} \in \mathbf{R}^{d \times m}$  es una matriz dada de rango  $\mathbf{m}$ , requerimos

$$\mathbf{B}'\Gamma = \begin{bmatrix} x & 0 & 0 & \cdots & 0 \\ x & x & 0 & \cdots & 0 \\ x & x & x & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ x & x & x & x & x \end{bmatrix}$$

(c) Diagonalidad de  $\Gamma'\Gamma$

Aquí pediremos que los elementos de la diagonal sean diferentes y ordenados en forma descendente.

Esto equivale a pedir, que el efecto de los factores sobre las  $\mathbf{x}$  este no correlacionado.

(d) Diagonalidad de  $\Gamma'\Psi^{-1}\Gamma$

Y también pediremos que los elementos de la diagonal sean diferentes y ordenados en forma descendente.

### Otras alternativas para eliminar la indeterminación de rotación

Las restricciones impuestas anteriormente son recursos matemáticos para lograr una solución única, pero no necesariamente obedecen a ninguna razón de interpretación práctica. Mencionaremos otras alternativas para restringir a  $\Gamma$ , que sí están fundadas en motivos prácticos.

#### ♦ Estructura Simple (Thurstone)

La idea es que cuando los factores tienen una interpretación útil, en general los tests no dependen de todos los factores. Luego entre todas las rotaciones de  $\Gamma$ , tendríamos que seleccionar una que tenga muchos "ceros". Precisando,  $\Gamma$  requeriría tener (Thurstone):

- (a) En cada fila al menos un cero.
- (b) En cada columna por lo menos  $\mathbf{m}$  ceros
- (c) En cada par de columnas, muchas filas con 1 o 2 ceros.

#### ♦ Elementos cero en posiciones predeterminadas

Por ejemplo puede ocurrir que el investigador sepa que algunos tests no dependen de algunos factores. Entonces, esto equivale a exigir que algunos elementos de  $\Gamma$  sean nulos.

## Identificación respecto de la Estructura.

### ■ Teorema de Condición Suficiente

$\forall$  fila de  $\Gamma$ , existen 2 submatrices disjuntas de rango  $\mathbf{m}$ , (que no la contienen)  $\Rightarrow (\Gamma; \Psi)$  **única s/r**

**Dem:** Si existiese  $(\mathbf{H}, \Delta)$  solución,  $\Gamma \Gamma' + \Psi = \Sigma = \mathbf{H} \mathbf{H}' + \Delta$

luego los elementos no diagonales de  $\Gamma \Gamma'$  y  $\mathbf{H} \mathbf{H}'$  serán iguales.

$$\text{Sean } \Gamma = \begin{bmatrix} \Gamma_1 \\ \gamma \\ \Gamma_2 \\ \Gamma_3 \end{bmatrix} \quad \text{con } \Gamma_1, \Gamma_2 \in \mathbb{R}^{m \times m} \text{ no singulares, y } \gamma \in \mathbb{R}^{1 \times m} \text{ } (\Gamma_3 \text{ es el resto})$$

Obviamente tendrá que ser  $2\mathbf{m} + 1 \leq \mathbf{d}$ .

Particiono a  $\mathbf{H}$  de la misma forma, con  $\mathbf{H}_1, \mathbf{H}_2, \mathbf{h}$  (y  $\mathbf{H}_3$ ).

$$\Gamma \Gamma' = \begin{bmatrix} \Gamma_1 \Gamma_1' & \Gamma_1 \gamma' & \Gamma_1 \Gamma_2' & \Gamma_1 \Gamma_3' \\ \gamma \Gamma_1' & \gamma \gamma' & \gamma \Gamma_2' & \gamma \Gamma_3' \\ \Gamma_2 \Gamma_1' & \Gamma_2 \gamma' & \Gamma_2 \Gamma_2' & \Gamma_2 \Gamma_3' \\ \Gamma_3 \Gamma_1' & \Gamma_3 \gamma' & \Gamma_3 \Gamma_2' & \Gamma_3 \Gamma_3' \end{bmatrix} \quad \mathbf{H} \mathbf{H}' = \begin{bmatrix} \mathbf{H}_1 \mathbf{H}_1' & \mathbf{H}_1 \mathbf{h}' & \mathbf{H}_1 \mathbf{H}_2' & \mathbf{H}_1 \mathbf{H}_3' \\ \mathbf{h} \mathbf{H}_1' & \mathbf{h} \mathbf{h}' & \mathbf{h} \mathbf{H}_2' & \mathbf{h} \mathbf{H}_3' \\ \mathbf{H}_2 \mathbf{H}_1' & \mathbf{H}_2 \mathbf{h}' & \mathbf{H}_2 \mathbf{H}_2' & \mathbf{H}_2 \mathbf{H}_3' \\ \mathbf{H}_3 \mathbf{H}_1' & \mathbf{H}_3 \mathbf{h}' & \mathbf{H}_3 \mathbf{H}_2' & \mathbf{H}_3 \mathbf{H}_3' \end{bmatrix}$$

Resultará:  $\Gamma_1 \gamma' = \mathbf{H}_1 \mathbf{h}' \quad \gamma \Gamma_2' = \mathbf{h} \mathbf{H}_2' \quad \Gamma_1 \Gamma_2' = \mathbf{H}_1 \mathbf{H}_2'$

pero  $\Gamma_1 \Gamma_2'$  no singular  $\rightarrow \text{rg}(\Gamma_1 \Gamma_2') = \mathbf{m} = \text{rg}(\mathbf{H}_1 \mathbf{H}_2')$

$\rightarrow \text{rg}(\mathbf{H}_1) = \mathbf{m} \rightarrow \text{rg}(\mathbf{H}) = \mathbf{m} \rightarrow \text{rg}(\mathbf{H} \mathbf{H}') = \mathbf{m}$

Luego son nulos los siguientes determinantes de orden  $m+1$

$$\begin{vmatrix} \Gamma_1 \gamma' & \Gamma_1 \Gamma_2' \\ \gamma \gamma' & \gamma \Gamma_2' \end{vmatrix} = \begin{vmatrix} \mathbf{H}_1 \mathbf{h}' & \mathbf{H}_1 \mathbf{H}_2' \\ \mathbf{h} \mathbf{h}' & \mathbf{h} \mathbf{H}_2' \end{vmatrix} = 0$$

$$\underbrace{g(\Gamma_1 \gamma'; \gamma \Gamma_2')}_{\uparrow} + (-1)^m \gamma \gamma' \left| \Gamma_1 \Gamma_2' \right| = \underbrace{g(\mathbf{H}_1 \mathbf{h}'; \mathbf{h} \mathbf{H}_2')}_{\uparrow} + (-1)^m \mathbf{h} \mathbf{h}' \left| \mathbf{H}_1 \mathbf{H}_2' \right|$$

↑————— son iguales —————↑

Y como  $\left| \Gamma_1 \Gamma_2' \right| = \left| \mathbf{H}_1 \mathbf{H}_2' \right|$  de aquí resulta que  $\gamma \gamma' = \mathbf{h} \mathbf{h}'$ .

Haciendo lo mismo para los otros elementos de la diagonal

$$\Rightarrow \Gamma \Gamma' = \mathbf{H} \mathbf{H}' \quad \Rightarrow (\Gamma; \Psi) \text{ **única s/r**}$$

**Observación:** Esta condición es demasiado fuerte. Wilson y Worcester mostraron un ejemplo

en que  $\mathbf{d} = 6$  y  $\mathbf{m} = 3$  con solución única. Además la condición  $2\mathbf{m} + 1 \leq \mathbf{d}$  es más exigente que  $\mathbf{C} \geq 0$ .

**Condiciones necesarias** (asumimos ahora que  $\Psi > \mathbf{0}$ )

Defino  $\mathbf{C}(\Gamma)$  una condición sobre  $\Gamma$ , que es cierta, si existe unicidad, o sea

$$\mathbf{C}(\Gamma) = \mathbf{V} \iff \text{Unicidad}$$

### ■ Lema Rotaciones

$$\boxed{[\mathbf{C}(\Gamma) = \mathbf{V} \iff \text{Unicidad}] \implies [\mathbf{C}(\Gamma\mathbf{L}) = \mathbf{V}, \forall \mathbf{L} \text{ ortogonal} \iff \text{Unicidad}]}$$

ya que si  $\mathbf{C}(\Gamma\mathbf{L}) = \mathbf{F} \rightarrow$  existen  $\Gamma\mathbf{L}(\Gamma\mathbf{L})' + \Psi = \mathbf{H}\mathbf{H}' + \Delta$  con  $\Gamma\mathbf{L}(\Gamma\mathbf{L})' \neq \mathbf{H}\mathbf{H}'$

existen  $\Gamma\Gamma' + \Psi = \mathbf{H}\mathbf{H}' + \Delta$  con  $\Gamma\Gamma' \neq \mathbf{H}\mathbf{H}'$

luego  $(\Gamma; \Psi)$  no sería única

En esencia este teorema dice que si la unicidad implica que  $\Gamma$  debe cumplir una condición, esta condición también debe cumplirla cualquier rotación de  $\Gamma$ .

### ■ Lema Columnas

Particionamos  $\Gamma$  por columnas así  $\Gamma = [\Gamma^*; \Gamma^{**}]$

$$\boxed{[\mathbf{C}(\Gamma) = \mathbf{V} \iff \text{Unicidad}] \implies [\mathbf{C}(\Gamma^*) = \mathbf{V} \iff \text{Unicidad}]}$$

ya que si  $\mathbf{C}(\Gamma^*) = \mathbf{F} \rightarrow$  existen  $\Gamma^*\Gamma^{**'} + \Psi = \mathbf{H}\mathbf{H}' + \Phi$  con  $\Gamma^*\Gamma^{**'} \neq \mathbf{H}\mathbf{H}'$

pero  $\Gamma^{**}\Gamma^{**'} = \Gamma^{**}\Gamma^{**'}$

$$\text{luego } \underbrace{\Gamma\Gamma'} + \Psi = \underbrace{[\mathbf{H}\Gamma^{**}][\mathbf{H}\Gamma^{**}]}' + \Phi$$

↑ ——— distintos ——— ↑

luego  $(\Gamma; \Psi)$  no sería única

Este teorema dice que si la unicidad implica que  $\Gamma$  debe cumplir una condición, esta condición también debe cumplirla cualquier grupo de columnas de  $\Gamma$ .

### ■ Lema Filas

Si  $\Gamma$  tiene filas todas nulas, particionamos por filas así

$$\Gamma = \begin{bmatrix} \Gamma^* \\ \mathbf{0} \end{bmatrix} \text{ y } \Psi = \begin{bmatrix} \Psi^* & \mathbf{0} \\ \mathbf{0} & \Psi^{**} \end{bmatrix} \text{ y } \Sigma = \begin{bmatrix} \Sigma^* & \mathbf{0} \\ \mathbf{0} & \Sigma^{**} \end{bmatrix}$$

$$\boxed{[\mathbf{C}(\Gamma) = \mathbf{V} \iff \text{Unicidad}] \implies [\mathbf{C}(\Gamma^*) = \mathbf{V} \iff \text{Unicidad}]}$$

ya que  $\Sigma = \Gamma\Gamma' + \Psi$  implica en este caso

$$\begin{bmatrix} \Sigma^* & \mathbf{0} \\ \mathbf{0} & \Sigma^{**} \end{bmatrix} = \begin{bmatrix} \Gamma^*\Gamma^{**'} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \Psi^* & \mathbf{0} \\ \mathbf{0} & \Psi^{**} \end{bmatrix}$$

luego  $\Sigma^* = \Gamma^*\Gamma^{**'} + \Psi^*$   
 $\Sigma^{**} = \Psi^{**}$

y solo la primera involucra a  $\Gamma^*$  y a  $\Psi^*$ .

Este teorema dice que si la unicidad implica que  $\Gamma$  debe cumplir una condición, esta condición también debe cumplirla las filas no nulas de  $\Gamma$ .

■ Teorema de NO Identificabilidad para  $m=1$   $d=2$

Si  $d=2$   $m=1 \Rightarrow (\Gamma, \Psi)$  no identificada

Si  $\Sigma = \Gamma\Gamma' + \Psi \Rightarrow \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \gamma_1^2 + \Psi_{11} & \gamma_1\gamma_2 \\ \gamma_1\gamma_2 & \gamma_2^2 + \Psi_{22} \end{bmatrix}$

Si  $\sigma_{12} = 0 \rightarrow$  algún  $\gamma_i$  es nulo por ejemplo  $\gamma_2 = 0$  y queda

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \gamma_1^2 + \Psi_{11} & 0 \\ 0 & \Psi_{22} \end{bmatrix}$$

aquí basta tomar  $\Psi_{22} = \sigma_2^2$  y elegir  $\Psi_{11}^* > 0$ ,  $\gamma_1^*$  cumpliendo  $\gamma_1^{*2} + \Psi_{11}^* = \sigma_1^2$

Si  $\sigma_{12} \neq 0 \rightarrow \gamma_1^*\gamma_2^* = \sigma_{12} \rightarrow \gamma_1^* = \sigma_{12}/\gamma_2^*$

Como  $\Psi_{22}^* > 0 \rightarrow \sigma_2^2 - \gamma_2^{*2} > 0 \rightarrow \gamma_2^{*2} < \sigma_2^2$

Como  $\Psi_{11}^* > 0 \rightarrow \sigma_1^2 - \gamma_1^{*2} = \sigma_1^2 - \sigma_{12}^2/\gamma_2^{*2} > 0 \rightarrow \sigma_{12}^2/\sigma_1^2 < \gamma_2^{*2}$

Luego debo elegir  $\gamma_2^*$  de manera que  $\sigma_{12}^2/\sigma_1^2 < \gamma_2^{*2} < \sigma_2^2$  y tomar

$\Psi_{22}^* = \sigma_2^2 - \gamma_2^{*2}$   $\gamma_1^* = \sigma_{12}/\gamma_2^*$   $\Psi_{11}^* = \sigma_1^2 - \gamma_1^{*2}$  y listo.

■ Teorema de Identificabilidad para  $m=1$

Al menos tres  $\gamma_i \neq 0 \Leftrightarrow (\Gamma; \Psi)$  única s/r

$\Rightarrow$  por el teorema de condición suficiente

$\Leftarrow$  pues si hay 2 o menos  $\gamma_i \neq 0$ , por el Lema Fila y el teorema de NO Identificabilidad para  $m=1$ ,  $(\Gamma; \Psi)$  no sería única.



■ Teorema de Condición Necesaria I

Al menos tres  $\gamma_i \neq 0$  en c/columna de  $\Gamma A$ ,  $\forall A$  no singular  $\Leftrightarrow (\Gamma; \Psi)$  única s/r

Si  $A = I$  sale del Lema Columna, y del teorema de Identificabilidad para  $m=1$ .

Luego el Lema de Rotaciones extiende su validez para  $A$  ortogonal.

Si  $A$  no es ortogonal, supongamos que en  $\Gamma A$  hay una columna con 2 o menos  $\gamma_i \neq 0$ , luego lo mismo vale para una matriz ortogonal, que tiene esa columna proporcional a la de  $A$  (absurdo).

■ Teorema de NO Identificabilidad para  $m=2$   $d=4$

Si  $d=4$   $m=2 \Rightarrow (\Gamma, \Psi)$  no identificada

■ Teorema de Identificabilidad para  $m=2$

$\forall$  fila de  $\Gamma$ , existen 2 submatrices disjuntas de rango 2, (que no la contienen)  $\Leftrightarrow (\Gamma; \Psi)$  única s/r

$\Rightarrow$  por el teorema de Condición Suficiente

$\Leftarrow$  supongamos que no existen, luego para la fila 1 se tendrá

$$\Gamma = \begin{bmatrix} \gamma_1 \\ \Gamma_2 \\ \Gamma_3 \end{bmatrix} \text{ con } \text{rg}(\Gamma_2) \leq 2 \quad \text{rg}(\Gamma_3) = 1, \text{ busco } L \text{ ortogonal con } \Gamma_3 L = \begin{bmatrix} v_1 & 0 \\ \vdots & 0 \\ v_{d-3} & 0 \end{bmatrix}$$

Luego

$$\Gamma L = \Gamma^* = \begin{bmatrix} \gamma_{11}^* & \gamma_{12}^* \\ \gamma_{21}^* & \gamma_{22}^* \\ \gamma_{31}^* & \gamma_{32}^* \\ v_1 & 0 \\ \vdots & 0 \\ v_{d-3} & 0 \end{bmatrix} \text{ por el teorema de Condición Necesaria I } \gamma_{12}^*, \gamma_{22}^*, \gamma_{32}^* \text{ son } \neq 0$$

Consideremos las dos submatrices de  $\mathbb{R}^{2 \times 2}$   $\begin{bmatrix} \gamma_{21}^* & \gamma_{22}^* \\ v_i & 0 \end{bmatrix}$  y  $\begin{bmatrix} \gamma_{31}^* & \gamma_{32}^* \\ v_j & 0 \end{bmatrix}$

Como supusimos que  $\nexists$  dos submatrices de rango = 2  $\rightarrow v_i=0$  o  $v_j=0$ .

Luego todos los  $v_i$  deben ser nulos salvo a lo sumo uno.

Por el Lema Fila y el teorema de No Identificabilidad para  $m=2$ ,  $(\Gamma; \Psi)$  no sería única.

## ■ Teorema de Condición Necesaria II

$\forall$  par de **columnas** de  $\Gamma A$ , con **A no singular**,  
 existen 2 submatrices disjuntas de rango **2**, (que no la contienen)  
 $\Leftrightarrow (\Gamma; \Psi)$  **única s/r**

Sale del Lema Rotaciones, Lema Columnas y del teorema de Identificabilidad para  $m = 2$ .

### *Que es más importante: Existencia o Identificabilidad?*

Desde cierto punto de vista es mas importante el problema de Identificación, ya que exige del investigador la especificación de ciertas características del modelo, y por tanto se debe conocer como hacer esto, y que implicancias tiene.

Por otro lado, la existencia se da o no, y es poco lo que se puede hacer.

### Identificación Local

Si  $\Gamma$  y  $\Psi$  cumplen  $\Sigma = \Gamma \Gamma' + \Psi$  y alguna otra restricción (por ejemplo  $\Gamma' \Psi^{-1} \Gamma$  diagonal) la pregunta es:

¿ si hago un cambio pequeño de  $\Gamma$  por  $\Gamma^* = \Gamma + \Delta \Gamma$ , y de  $\Psi$  por  $\Psi^* = \Psi + \Delta \Psi$ , se mantendrá  $\Gamma^* \Gamma^{*'} + \Psi^* = \Sigma$  ?

Si esto ocurre se dice que  $\Gamma$  y  $\Psi$  están localmente identificados.

## 4) DETERMINACION DE LA ESTRUCTURA

(3er problema de la población)

Dada  $\Sigma$ , si se cumplen las hipótesis del teorema de Condición Suficiente, las comunalidades se pueden obtener siguiendo el método de este teorema. Esto permite conocer  $\Gamma \Gamma' = C$  y  $\Psi$ .

Si llamo  $\Gamma$  ( $\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(m)}$ )

$$C = \gamma^{(1)}\gamma^{(1)'} + \gamma^{(2)}\gamma^{(2)'} + \dots + \gamma^{(m)}\gamma^{(m)'}$$

Hay que hallar  $\Gamma$  y esto depende de las condiciones que impusimos para lograr identificación.

(a) Forma triangular de  $\Gamma$

Supongo

$$\Gamma = \begin{bmatrix} \gamma_{11} & 0 & 0 & \cdots & 0 \\ \gamma_{21} & \gamma_{22} & 0 & \cdots & 0 \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \gamma_{m3} & \cdots & \gamma_{mm} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{d1} & \gamma_{d2} & \gamma_{d3} & \cdots & \gamma_{dm} \end{bmatrix} \quad \text{luego}$$

$$\Gamma\Gamma' = \begin{bmatrix} \gamma_{11}^2 & \cdot & \cdot & \cdots & \cdot \\ \gamma_{21}\gamma_{11} & \cdot & \cdot & \cdots & \cdot \\ \gamma_{31}\gamma_{11} & \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{d1}\gamma_{11} & \cdot & \cdot & \cdots & \cdot \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1d} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2d} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3d} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{d1} & c_{d2} & c_{d3} & \cdots & c_{dd} \end{bmatrix} = \mathbf{C}$$

De  $\gamma_{11}^2 = c_{11} \rightarrow$  sale  $\gamma_{11}$

Luego con el resto de las primeras columnas

$$\gamma_{i1}\gamma_{11} = c_{i1} \text{ para } i = 2 \dots d, \text{ salen los } \gamma_{i1}$$

Ahora calculo

$$\mathbf{C}^{(-1)} = \mathbf{C} - \gamma^{(1)}\gamma^{(1)'} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & x & x & \cdots & x \\ 0 & x & x & \cdots & x \\ 0 & \vdots & \vdots & & \vdots \\ 0 & x & x & \cdots & x \end{bmatrix} \text{ (con las x, no necesariamente nulos)}$$

$$\Gamma^{(-1)}\Gamma^{(-1)'} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \gamma_{22}^2 & \cdot & \cdots & \cdot \\ 0 & \gamma_{32}\gamma_{22} & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \gamma_{d2}\gamma_{22} & \cdot & \cdots & \cdot \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & c_{22}^* & c_{23}^* & \cdots & c_{2d}^* \\ 0 & c_{32}^* & c_{33}^* & \cdots & c_{3d}^* \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & c_{d2}^* & c_{d3}^* & \cdots & c_{dd}^* \end{bmatrix} = \mathbf{C}^{(-1)}$$

De  $\gamma_{22}^2 = c_{22}^* \rightarrow$  sale  $\gamma_{22}$

Luego con el resto de las primeras columnas

$$\gamma_{i2}\gamma_{22} = c_{i2}^* \text{ para } i = 3 \dots d, \text{ salen los } \gamma_{i2}$$

Y siguiendo así se hallan todos.

(b) Condición general de triangularidad

Aquí supongo que conozco  $\mathbf{B} \in \mathbf{R}^{d \times m}$  y quiero que

$$\mathbf{F} = \mathbf{B}'\Gamma = \begin{bmatrix} f_{11} & 0 & 0 & \cdots & 0 \\ f_{21} & f_{22} & 0 & \cdots & 0 \\ f_{31} & f_{32} & f_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & f_{m3} & \cdots & f_{mm} \end{bmatrix} = (f^{(1)}, f^{(2)}, f^{(3)}, \dots, f^{(m)})$$

Con  $\mathbf{B} = (b^{(1)}, b^{(2)}, b^{(3)}, \dots, b^{(m)})$  y además se tiene

Virgilio L. Foglia (2003)

$$\mathbf{CB} = \Gamma\Gamma'\mathbf{B} = \Gamma(\mathbf{B}'\Gamma)' = \Gamma\mathbf{F}' = \gamma^{(1)}f^{(1)'} + \gamma^{(2)}f^{(2)'} + \gamma^{(3)}f^{(3)'} + \dots + \gamma^{(m)}f^{(m)'}$$

$$\mathbf{B}'\mathbf{CB} = \mathbf{B}'\Gamma\Gamma'\mathbf{B} = \mathbf{B}'\Gamma(\mathbf{B}'\Gamma)' = \mathbf{F}\mathbf{F}' = f^{(1)}f^{(1)'} + f^{(2)}f^{(2)'} + f^{(3)}f^{(3)'} + \dots + f^{(m)}f^{(m)'}$$

Del primer elemento de  $\mathbf{B}'\mathbf{CB} = f_{11}^2 \rightarrow$  sale  $f_{11}$

La primera columna de  $\mathbf{CB}$  es 
$$\begin{bmatrix} \gamma_{11}f_{11} \\ \gamma_{21}f_{11} \\ \vdots \\ \gamma_{d1}f_{11} \end{bmatrix} \rightarrow$$
 sale  $\gamma^{(1)}$

El elemento (1,2) de  $\mathbf{B}'\mathbf{CB}$  es  $f_{11}f_{21} \rightarrow$  sale  $f_{21}$

El elemento (2,2) de  $\mathbf{B}'\mathbf{CB}$  es  $f_{21}^2 + f_{22}^2 \rightarrow$  sale  $f_{22}$

La segunda columna de  $\mathbf{CB}$  es  $\gamma^{(1)}f_{21} + \gamma^{(2)}f_{22} \rightarrow$  sale  $\gamma^{(2)}$

Y se sigue así.

(c) Diagonalidad de  $\Gamma'\Gamma$

Como  $\text{rg}(\mathbf{C}) = \mathbf{m}$ , tomo los  $\mathbf{m}$  autovalores  $\neq 0$  ordenados en forma decreciente

$$\mathbf{C} = \mathbf{T}\Lambda\mathbf{T}' = \mathbf{T}_m\Lambda_m\mathbf{T}_m' = \mathbf{T}_m\Lambda_m^{1/2}\Lambda_m^{1/2}\mathbf{T}_m' = \mathbf{T}_m\Lambda_m^{1/2}(\mathbf{T}_m\Lambda_m^{1/2})' = \Gamma\Gamma'$$

Luego  $\Gamma = \mathbf{T}_m\Lambda_m^{1/2} \in \mathbb{R}^{d \times m}$  es único

$$\text{con } \Gamma'\Gamma = (\mathbf{T}_m\Lambda_m^{1/2})'\mathbf{T}_m\Lambda_m^{1/2} = \Lambda_m^{1/2}'\mathbf{T}_m'\mathbf{T}_m\Lambda_m^{1/2} = \Lambda_m \text{ (diagonal)}$$

(d) Diagonalidad de  $\Gamma'\Psi^{-1}\Gamma$

Quiero  $\mathbf{C} = \Gamma\Gamma'$  con la restricción  $\Gamma'\Psi^{-1}\Gamma = \mathbf{D}$  (diagonal)

Luego  $\mathbf{C}\Psi^{-1}\Gamma = \Gamma\underbrace{\Gamma'\Psi^{-1}\Gamma}_{\mathbf{D}} = \Gamma\mathbf{D} \rightarrow \mathbf{D}$  tiene los primeros  $\mathbf{m}$  autovalores de  $\mathbf{C}\Psi^{-1}$

Y como debe cumplirse  $\Gamma'\Psi^{-1}\Gamma = \mathbf{D} \rightarrow \mathbf{D}^{-1/2}\Gamma'\Psi^{-1/2}\underbrace{\Psi^{-1/2}\Gamma\mathbf{D}^{-1/2}}_{\mathbf{I}_m} = \mathbf{I}_m$

Donde  $\Psi^{-1/2}\Gamma\mathbf{D}^{-1/2} = \mathbf{T}$  que tiene los primeros  $\mathbf{m}$  autovectores de norma 1 de  $\mathbf{C}\Psi^{-1}$

Luego sale  $\Gamma = \Psi^{1/2}\mathbf{T}\mathbf{D}^{1/2}$

## 5) METODOS DE ESTIMACION

### (a) Máxima verosimilitud

Supongo  $\mathbf{x} \sim N_d(\mathbf{0}; \Sigma) \Rightarrow \mathbf{nS} \sim W_d(\mathbf{n}-1; \Sigma)$

y el modelo  $\Sigma = \Gamma\Gamma' + \Psi$

luego 
$$\ln L(\Gamma; \Psi) = \mathbf{c} - \frac{\mathbf{n}}{2} \{ \ln|\Sigma| + \text{Tr}(\Sigma^{-1}\mathbf{S}) \}$$

o también

$$\ln \mathbf{L}(\Gamma; \Psi) = \mathbf{c} - \frac{n}{2} \{ \ln |\Gamma \Gamma' + \Psi| + \text{Tr}((\Gamma \Gamma' + \Psi)^{-1} \mathbf{S}) \}$$

Impongo la condición  $\Gamma' \Psi^{-1} \Gamma = \mathbf{D}$  sea diagonal (unicidad por rotaciones)

Luego tenemos que maximizar  $\ln \mathbf{L}(\Gamma; \Psi)$  respecto de  $\Gamma$  y  $\Psi$  (sujeto a la restricción que  $\Gamma' \Psi^{-1} \Gamma$  sea diagonal).

Usando  $\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = \mathbf{X}^{-1}$ ,  $\frac{\partial \text{Tr}(\mathbf{X}\mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}'$ ,  $\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{X}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{X}} \mathbf{A}^{-1}$ ,  $\frac{\partial \mathbf{X}\mathbf{X}'}{\partial \mathbf{X}} = 2\mathbf{X}$

$$\frac{\partial \mathbf{L}(\Gamma; \Psi)}{\partial \Gamma} = \Sigma^{-1} 2\Gamma - \Sigma^{-1} \mathbf{S} \Sigma^{-1} 2\Gamma = 2\Sigma^{-1}(\Sigma - \mathbf{S})\Sigma^{-1} \Gamma = \mathbf{0}$$

$$\frac{\partial \mathbf{L}(\Gamma; \Psi)}{\partial \Psi} = \text{diag}(\Sigma^{-1} - \Sigma^{-1} \mathbf{S} \Sigma^{-1}) = \mathbf{0}$$

De la primera resulta  $(\Sigma - \mathbf{S})\Sigma^{-1} \Gamma = \mathbf{0}$   
 $\Gamma - \mathbf{S}\Sigma^{-1} \Gamma = \mathbf{0}$

y ahora usando que  $\Sigma^{-1} = (\Gamma \Gamma' + \Psi)^{-1} = \Psi^{-1} - \Psi^{-1} \Gamma (\mathbf{I}_m + \Gamma' \Psi^{-1} \Gamma)^{-1} \Gamma' \Psi^{-1}$ , y reemplazando

$$\Gamma - \mathbf{S} \left\{ \Psi^{-1} - \Psi^{-1} \Gamma (\mathbf{I}_m + \Gamma' \Psi^{-1} \Gamma)^{-1} \Gamma' \Psi^{-1} \right\} \Gamma = \mathbf{0}$$

$$\Gamma - \mathbf{S} \left\{ \Psi^{-1} \Gamma - \Psi^{-1} \Gamma (\mathbf{I}_m + \Gamma' \Psi^{-1} \Gamma)^{-1} \Gamma' \Psi^{-1} \Gamma \right\} = \mathbf{0}$$

incorporando la restricción  $\Gamma' \Psi^{-1} \Gamma = \mathbf{D}$

$$\Gamma - \mathbf{S} \left\{ \Psi^{-1} \Gamma - \Psi^{-1} \Gamma (\mathbf{I}_m + \mathbf{D})^{-1} \mathbf{D} \right\} = \mathbf{0}$$

$$\Gamma - \mathbf{S} \left\{ \Psi^{-1} \Gamma - \Psi^{-1} \Gamma (\mathbf{I}_m + \mathbf{D})^{-1} (\mathbf{D} + \mathbf{I}_m - \mathbf{I}_m) \right\} = \mathbf{0}$$

$$\Gamma - \mathbf{S} \left\{ \Psi^{-1} \Gamma - \Psi^{-1} \Gamma (\mathbf{I}_m - (\mathbf{I}_m + \mathbf{D})^{-1}) \right\} = \mathbf{0}$$

$$\Gamma - \mathbf{S} \left\{ \Psi^{-1} \Gamma - \Psi^{-1} \Gamma + \Psi^{-1} \Gamma (\mathbf{I}_m + \mathbf{D})^{-1} \right\} = \mathbf{0}$$

$$\Gamma - \mathbf{S} \Psi^{-1} \Gamma (\mathbf{I}_m + \mathbf{D})^{-1} = \mathbf{0}$$

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

$$\mathbf{S} \Psi^{-1} \Gamma = \Gamma (\mathbf{I}_m + \mathbf{D}) \quad (\text{primera ecuación de Lawley})$$

Para llegar a la tercera ecuación de Lawley parto de la primera derivada parcial de la verosimilitud y queda

$$(\mathbf{S} - \Sigma)\Sigma^{-1} \Gamma = \mathbf{0}$$

$$(\mathbf{S} - \Sigma)\Sigma^{-1} \Gamma \Gamma' = \mathbf{0}$$

$$(\mathbf{S} - \Sigma)\Sigma^{-1} (\Sigma - \Psi) = \mathbf{0}$$

$$(\mathbf{S} - \Sigma)(\mathbf{I}_d - \Sigma^{-1} \Psi) = \mathbf{0}$$

$$(\mathbf{S} - \Sigma) = (\mathbf{S} - \Sigma) \Sigma^{-1} \Psi \quad (\text{ecuación A})$$

$$\Sigma^{-1} (\mathbf{S} - \Sigma) = \Sigma^{-1} (\mathbf{S} - \Sigma) \Sigma^{-1} \Psi$$

transponiendo  $(\mathbf{S} - \Sigma)\Sigma^{-1} = \Psi \Sigma^{-1} (\mathbf{S} - \Sigma) \Sigma^{-1} \quad (\text{ecuación B})$

de la segunda derivada parcial de la verosimilitud

$$\text{diag}(\Sigma^{-1} - \Sigma^{-1} \mathbf{S} \Sigma^{-1}) = \mathbf{0}$$

$$\text{diag}(\Sigma^{-1} (\mathbf{S} - \Sigma) \Sigma^{-1}) = \mathbf{0}$$

por la ecuación B, y dado que  $\Psi$  es diagonal

$$\text{diag} \left\{ (\mathbf{S} - \Sigma) \Sigma^{-1} \right\} = \text{diag} \left\{ \Psi \Sigma^{-1} (\mathbf{S} - \Sigma) \Sigma^{-1} \right\} = \text{diag}(\Sigma^{-1} (\mathbf{S} - \Sigma) \Sigma^{-1}) = \mathbf{0}$$

o sea

$$\text{diag} \left\{ (\mathbf{S} - \Sigma) \Sigma^{-1} \right\} = \mathbf{0}$$

Virgilio L. Foglia (2003)

pero por la ecuación A, y usando que  $\Psi$  es diagonal

$$\mathbf{diag}\{(\mathbf{S} - \Sigma)\} = \mathbf{diag}\{(\mathbf{S} - \Sigma)\Sigma^{-1}\Psi\} = \mathbf{diag}\{(\mathbf{S} - \Sigma)\Sigma^{-1}\} = \mathbf{0}$$

o sea

$$\mathbf{diag}(\mathbf{S} - \Sigma) = \mathbf{0}$$

o sea

$$\mathbf{diag}(\mathbf{S} - \Gamma\Gamma' - \Psi) = \mathbf{0}$$

y resulta la tercera ecuación de Lawley

$$\Psi = \mathbf{diag}(\mathbf{S} - \Gamma\Gamma')$$

Luego en definitiva resultan las ecuaciones (Lawley)

$$\begin{aligned} \mathbf{S}\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\Gamma}(\mathbf{I}_m + \hat{\mathbf{D}}) \\ \hat{\Gamma}'\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\mathbf{D}} \text{ (diagonal)} \\ \hat{\Psi} &= \mathbf{diag}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}') \end{aligned}$$

Transformando la primera queda:

$$\mathbf{S}\hat{\Psi}^{-1}\hat{\Gamma} - \hat{\Gamma} = \hat{\Gamma}\hat{\mathbf{D}}$$

$$\mathbf{S}\hat{\Psi}^{-1}\hat{\Gamma} - \hat{\Psi}\hat{\Psi}^{-1}\hat{\Gamma} = \hat{\Gamma}\hat{\mathbf{D}}$$

$$(\mathbf{S} - \hat{\Psi})\hat{\Psi}^{-1}\hat{\Gamma} = \hat{\Gamma}\hat{\mathbf{D}}$$

Que se pueden expresar también así:

$$\begin{aligned} (\mathbf{S} - \hat{\Psi})\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\Gamma}\hat{\mathbf{D}} \\ \hat{\Gamma}'\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\mathbf{D}} \text{ (diagonal)} \\ \hat{\Psi} &= \mathbf{diag}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}') \end{aligned}$$

Estas ecuaciones hay que resolverlas por un algoritmo iterativo, pero antes necesitamos una solución inicial para  $\Psi$ .

## Estimación inicial de $\Psi$

En  $\mathbf{x} = \Gamma\mathbf{f} + \epsilon$  tomemos por ejemplo la primera fila

$$\mathbf{x}_1 = \gamma_{11}\mathbf{f}_1 + \gamma_{12}\mathbf{f}_2 + \dots + \gamma_{1m}\mathbf{f}_m + \epsilon_1$$

si conociese el valor de los factores, podría estimar con un modelo lineal  $\widehat{\text{var}}(\epsilon_1) = \widehat{\Psi}_1$  y listo.

Que pasa si planteo el modelo (esto sí es posible)

$$\boxed{\mathbf{x}_1 = \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + \dots + \beta_d\mathbf{x}_d + \delta_1}$$
 de aquí sale  $\widehat{\text{var}}(\delta_1) = \widehat{\sigma}_R^2$

Pero sabiendo que para  $i = 2 \dots d$   $\mathbf{x}_i = \gamma_{i1}\mathbf{f}_1 + \gamma_{i2}\mathbf{f}_2 + \dots + \gamma_{im}\mathbf{f}_m + \epsilon_i$  reemplazando

Virgilio L. Foglia (2003)

$$\mathbf{x}_1 = \gamma_{11}^* \mathbf{f}_1 + \gamma_{12}^* \mathbf{f}_2 + \dots + \gamma_{1m}^* \mathbf{f}_m + \underbrace{(\beta_2 \varepsilon_2 + \beta_3 \varepsilon_3 + \dots + \beta_d \varepsilon_d + \delta_1)}_{\widehat{\Psi}_1}$$

Notar entonces que al hacer la regresión de  $\mathbf{x}_1$  sobre  $\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_d$ , la varianza residual que se obtiene,  $\hat{\sigma}_R^2$ , no es igual a  $\widehat{\Psi}_1$ , sino a la varianza de  $\beta_2 \varepsilon_2 + \beta_3 \varepsilon_3 + \dots + \beta_d \varepsilon_d + \delta_1$ , que incluye a  $\widehat{\Psi}_1$ . Sin embargo, como se puede demostrar que  $\hat{\sigma}_R^2$  es una cota superior de  $\widehat{\Psi}_1$ , se toma como aproximación inicial

$$\widehat{\Psi}_1 = \hat{\sigma}_R^2$$

También se demuestra que si  $\mathbf{S}$  es la matriz de covarianza de las observaciones, y  $s^{ii} = [\mathbf{S}^{-1}]_{ii}$

resulta  $\boxed{\widehat{\Psi}_i = 1/s^{ii}} \rightarrow \boxed{\widehat{\Psi} = \mathbf{diag}^{-1}(\mathbf{S}^{-1})}$

Jöreskog propone  $\widehat{\Psi}_i = (1 - \frac{1}{2} \frac{m}{d}) 1/s^{ii}$  . o sea  $\widehat{\Psi} = (1 - \frac{1}{2} \frac{m}{d}) \mathbf{diag}^{-1}(\mathbf{S}^{-1})$

### Estimación de $\Gamma$ , conocida una estimación de $\Psi$

Parto de  $\widehat{\mathbf{S}} \widehat{\Psi}^{-1} \widehat{\Gamma} = \widehat{\Gamma} (\mathbf{I}_m + \widehat{\mathbf{D}})$  (primera ecuación de MV)

$$\widehat{\Psi}^{-1/2} \widehat{\mathbf{S}} \widehat{\Psi}^{-1/2} \underbrace{\widehat{\Psi}^{-1/2} \widehat{\Gamma}}_{\widehat{\Gamma}} = \underbrace{\widehat{\Psi}^{-1/2} \widehat{\Gamma}}_{\widehat{\Gamma}} (\mathbf{I}_m + \widehat{\mathbf{D}})$$

Pero entonces  $(\mathbf{I}_m + \widehat{\mathbf{D}})$  tiene autovalores de  $\widehat{\Psi}^{-1/2} \widehat{\mathbf{S}} \widehat{\Psi}^{-1/2}$ .

Tomo los primeros  $m$  autovalores, y llamo  $\Lambda_m = (\mathbf{I}_m + \widehat{\mathbf{D}})$ .

Además  $\widehat{\Psi}^{-1/2} \widehat{\Gamma}$  tendrá los correspondientes autovectores de norma 1, salvo un múltiplo escalar, luego

$$\widehat{\Psi}^{-1/2} \widehat{\Gamma} = \mathbf{T}_m \mathbf{J} \quad (\text{con } \mathbf{J} \text{ diagonal})$$

$$\widehat{\Gamma} = \widehat{\Psi}^{1/2} \mathbf{T}_m \mathbf{J}$$

Pero debe cumplirse la segunda ecuación de MV

$$\widehat{\Gamma}' \widehat{\Psi}^{-1} \widehat{\Gamma} = \mathbf{J}' \mathbf{T}_m' \widehat{\Psi}^{1/2} \widehat{\Psi}^{-1} \widehat{\Psi}^{1/2} \mathbf{T}_m \mathbf{J} = \mathbf{J}' \widehat{\mathbf{D}} = \Lambda_m - \mathbf{I}_m$$

luego  $\mathbf{J} = (\Lambda_m - \mathbf{I}_m)^{1/2}$

y quedará  $\boxed{\widehat{\Gamma} = \widehat{\Psi}^{1/2} \mathbf{T}_m (\Lambda_m - \mathbf{I}_m)^{1/2}}$

Resta verificar la tercera ecuación de MV.

Virgilio L. Foglia (2003)

$$\text{Como } \widehat{\Psi}^{-1/2} \mathbf{S} \widehat{\Psi}^{-1/2} = \mathbf{T} \begin{bmatrix} \Lambda_m & \mathbf{0} \\ \mathbf{0} & \Lambda_{d-m} \end{bmatrix} \mathbf{T}' = \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}'$$

$$\begin{aligned} \text{y } \widehat{\Psi}^{-1/2} \widehat{\Gamma\Gamma}' \widehat{\Psi}^{-1/2} &= \widehat{\Psi}^{-1/2} \widehat{\Psi}^{1/2} \mathbf{T}_m (\Lambda_m - \mathbf{I}_m) \mathbf{T}_m' \widehat{\Psi}^{1/2} \widehat{\Psi}^{-1/2} \\ &= \mathbf{T}_m (\Lambda_m - \mathbf{I}_m) \mathbf{T}_m' \\ &= \mathbf{T}_m \Lambda_m \mathbf{T}_m' - \mathbf{T}_m \mathbf{I}_m \mathbf{T}_m' \end{aligned}$$

$$\text{y } \widehat{\Psi}^{-1/2} \widehat{\Psi} \widehat{\Psi}^{-1/2} = \mathbf{I}_d = \mathbf{T}_m \mathbf{I}_m \mathbf{T}_m' + \mathbf{T}_{d-m} \mathbf{I}_{d-m} \mathbf{T}_{d-m}'$$

juntando las tres resulta

$$\widehat{\Psi}^{-1/2} (\mathbf{S} - \widehat{\Gamma\Gamma}' - \widehat{\Psi}) \widehat{\Psi}^{-1/2} = \mathbf{T}_{d-m} (\Lambda_{d-m} - \mathbf{I}_{d-m}) \mathbf{T}_{d-m}'$$

$$\widehat{\Psi}^{-1/2} (\mathbf{S} - \widehat{\Gamma\Gamma}' - \widehat{\Psi}) \widehat{\Psi}^{-1/2} = \mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}' - \mathbf{I}_d$$

$$\mathbf{S} - \widehat{\Gamma\Gamma}' - \widehat{\Psi} = \widehat{\Psi}^{1/2} \mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}' \widehat{\Psi}^{1/2} - \widehat{\Psi}$$

$$\mathbf{S} - \widehat{\Gamma\Gamma}' = \widehat{\Psi}^{1/2} \mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}' \widehat{\Psi}^{1/2}$$

### Primer algoritmo iterativo (Jöreskog)

Aunque ineficiente, Jöreskog propuso inicialmente como algoritmo la maximización directa de  $\ln \mathbf{L}(\Gamma; \Psi)$  así:

1°) Hacemos una estimación inicial  $\widehat{\Psi}^0 = (1 - \frac{1}{2} \frac{m}{d}) \mathbf{diag}^{-1}(\mathbf{S}^{-1})$

2°) Con  $\widehat{\Psi}^0$  estimo  $\widehat{\Gamma}^0 = \widehat{\Psi}^{0/2} \mathbf{T}_m^0 (\Lambda_m^0 - \mathbf{I}_m)^{1/2}$

3°) Se maximiza  $\ln \mathbf{L}(\widehat{\Gamma}^0; \Psi)$  respecto de  $\Psi$ , usando Fletcher & Powell, y se obtiene  $\widehat{\Psi}^1$ , y de aquí sale  $\widehat{\Gamma}^1 = \widehat{\Psi}^{1/2} \mathbf{T}_m^1 (\Lambda_m^1 - \mathbf{I}_m)^{1/2}$ , y luego  $\widehat{\Psi}^2$ , y se sigue iterando así.

### Transformación de la verosimilitud

En lugar de maximizar  $\ln \mathbf{L}(\Gamma; \Psi) = \mathbf{c} - \frac{n}{2} \{ \ln |\Gamma\Gamma' + \Psi| + \mathbf{Tr}((\Gamma\Gamma' + \Psi)^{-1} \mathbf{S}) \}$

minimizaremos  $\ln |\Gamma\Gamma' + \Psi| + \mathbf{Tr}((\Gamma\Gamma' + \Psi)^{-1} \mathbf{S}) = .\ln |\Sigma| + \mathbf{Tr}(\Sigma^{-1} \mathbf{S})$

En cada paso  $\widehat{\Sigma} = \widehat{\Gamma\Gamma}' + \widehat{\Psi} = \widehat{\Psi}^{1/2} \mathbf{T}_m (\Lambda_m - \mathbf{I}_m) \mathbf{T}_m' \widehat{\Psi}^{1/2} + \widehat{\Psi}$

$$\begin{aligned} \widehat{\Psi}^{-1/2} \widehat{\Sigma} \widehat{\Psi}^{-1/2} &= \mathbf{T}_m (\Lambda_m - \mathbf{I}_m) \mathbf{T}_m' + \mathbf{I}_d \\ &= \mathbf{T}_m (\Lambda_m - \mathbf{I}_m) \mathbf{T}_m' + \mathbf{T}_m \mathbf{I}_m \mathbf{T}_m' + \mathbf{T}_{d-m} \mathbf{I}_{d-m} \mathbf{T}_{d-m}' \\ &= \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \mathbf{T}_{d-m} \mathbf{I}_{d-m} \mathbf{T}_{d-m}' \end{aligned}$$

O sea  $\widehat{\Psi}^{-1/2} \widehat{\Sigma} \widehat{\Psi}^{-1/2} = \mathbf{T} \begin{bmatrix} \Lambda_m & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d-m} \end{bmatrix} \mathbf{T}' \quad (= \mathbf{T} \Lambda^* \mathbf{T}')$

Luego  $\left| \widehat{\Psi}^{-1/2} \widehat{\Sigma} \widehat{\Psi}^{-1/2} \right| = |\Lambda_m| \rightarrow \left| \widehat{\Sigma} \right| = |\Lambda_m| \left| \widehat{\Psi} \right|$



$$\boxed{|\hat{\Sigma}| = \prod_{j=1}^m \lambda_j \prod_{j=1}^d \psi_j}$$

Invirtiendo  $\hat{\Psi}^{1/2} \hat{\Sigma}^{-1} \hat{\Psi}^{1/2} = \mathbf{T} \Lambda^* \mathbf{T}'$ ,

$$\hat{\Sigma}^{-1} = \hat{\Psi}^{-1/2} \mathbf{T} \Lambda^{*-1} \mathbf{T}' \hat{\Psi}^{-1/2}$$

$$\hat{\Sigma}^{-1} \mathbf{S} = \hat{\Psi}^{-1/2} \mathbf{T} \Lambda^{*-1} \mathbf{T}' \hat{\Psi}^{-1/2} \mathbf{S}$$

sabemos que  $\underbrace{\hat{\Psi}^{-1/2} \mathbf{S} \hat{\Psi}^{-1/2}} \mathbf{T} = \mathbf{T} \Lambda \rightarrow \hat{\Psi}^{-1/2} \mathbf{S} = \mathbf{T} \Lambda \mathbf{T}' \hat{\Psi}^{1/2}$

luego  $\hat{\Sigma}^{-1} \mathbf{S} = \hat{\Psi}^{-1/2} \mathbf{T} \Lambda^{*-1} \mathbf{T}' \mathbf{T} \Lambda \mathbf{T}' \hat{\Psi}^{1/2}$

$$= \hat{\Psi}^{-1/2} \mathbf{T} \Lambda^{*-1} \Lambda \mathbf{T}' \hat{\Psi}^{1/2}$$

$$\mathbf{Tr}(\hat{\Sigma}^{-1} \mathbf{S}) = \mathbf{Tr}(\Lambda^{*-1} \Lambda \underbrace{\mathbf{T}' \hat{\Psi}^{1/2} \hat{\Psi}^{-1/2} \mathbf{T})$$

$$= \mathbf{Tr}(\Lambda^{*-1} \Lambda)$$

$$= \mathbf{Tr} \left( \begin{bmatrix} \Lambda_m^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d-m} \end{bmatrix} \begin{bmatrix} \Lambda_m & \mathbf{0} \\ \mathbf{0} & \Lambda_{d-m} \end{bmatrix} \right)$$

$$= \mathbf{Tr} \left( \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \Lambda_{d-m} \end{bmatrix} \right)$$

Luego

$$\boxed{\mathbf{Tr}(\hat{\Sigma}^{-1} \mathbf{S}) = m + \sum_{j=m+1}^d \lambda_j}$$

Finalmente, la expresión a minimizar es

$$\boxed{\ln|\Sigma| + \mathbf{Tr}(\Sigma^{-1} \mathbf{S}) = \sum_{j=1}^m \ln \lambda_j + \sum_{j=1}^d \ln \psi_j + m + \sum_{j=m+1}^d \lambda_j}$$

## Segundo algoritmo iterativo (Jöreskog 1967)

1°) Hacemos una estimación inicial  $\hat{\Psi}^0 = (1 - \frac{1}{2} \frac{m}{d}) \mathbf{diag}^{-1}(\mathbf{S}^{-1})$

2°) Con  $\hat{\Psi}^0$   $\hat{\Sigma}^{-1/2} \mathbf{S} \hat{\Sigma}^{-1/2}$  calculo los primeros  $m$  autovalores y obtengo  $\Lambda_m^0$

se reemplaza en  $\sum_{j=1}^m \ln \lambda_j^0 + \sum_{j=1}^d \ln \psi_j^0 + m + \sum_{j=m+1}^d \lambda_j^0$

3°) Se va iterando con otro  $\hat{\Psi}^i$ , calculando los

primeros  $m$  autovalores de  $\hat{\Psi}^i$   $\hat{\Sigma}^{-1/2} \mathbf{S} \hat{\Sigma}^{-1/2}$ , o sea  $\Lambda_m^i$ , y reemplazando

en  $\sum_{j=1}^m \ln \lambda_j^i + \sum_{j=1}^d \ln \psi_j^i + m + \sum_{j=m+1}^d \lambda_j^i$  hasta lograr convergencia en  $\hat{\Psi}^c$ .

Virgilio L. Foglia (2003)

4°) Finalmente con el  $\widehat{\Psi}^c$  estimado se calcula  $\widehat{\Gamma}^c = \widehat{\Psi}^c{}^{1/2} \mathbf{T}_m^c (\Lambda_m^c - \mathbf{I}_m)^{1/2}$

**Problema:** Suele suceder que en el proceso iterativo se llegue a un máximo en que  $\Psi < \mathbf{0}$ , a esto se le llama un caso "Heywood".

**Solución:** La maximización se conduce en una región en que  $\Psi_i \geq \varepsilon$  (pequeño  $\forall_i$ ) y se considera apropiada la solución si  $\Psi_i > \varepsilon$  para todo  $i$ .

Observación: Notar que la imposición  $\Gamma' \widehat{\Psi}^{-1} \Gamma = \mathbf{D}$  (**diagonal**) implica que el primer factor hará la máxima contribución a la varianza de las  $\mathbf{x}$ 's, el segundo hará la máxima contribución, sujeto a ser incorrelacionado con el primero y así siguiendo (Lawley & Maxwell, 1971, Cap 4).

### (b) Método de las componentes principales

Aquí supongo que  $\Psi = \sigma^2 \mathbf{I}_d$

Luego de

$$\begin{aligned} \mathbf{S}\Psi^{-1}\Gamma &= \Gamma(\mathbf{I}_m + \mathbf{D}) & \mathbf{S}\Gamma &= \Gamma\sigma^2(\mathbf{I}_m + \mathbf{D}) \\ \Gamma'\Psi^{-1}\Gamma &= \mathbf{D} \text{ (diagonal)} & \Rightarrow & \Gamma'\Gamma = \sigma^2\mathbf{D} \text{ (diagonal)} \\ \Psi &= \text{diag}(\mathbf{S} - \Gamma\Gamma') & & \sigma^2\mathbf{I}_d = \text{diag}(\mathbf{S} - \Gamma\Gamma') \end{aligned}$$

Luego  $\Lambda_m = \sigma^2(\mathbf{I}_m + \mathbf{D})$  tiene los primeros  $\mathbf{m}$  autovalores de  $\mathbf{S}$

Queda

$$\begin{aligned} \mathbf{S}\Gamma &= \Gamma\Lambda_m \\ \Gamma'\Gamma &= \Lambda_m - \sigma^2\mathbf{I}_m \text{ (diagonal)} \\ \sigma^2\mathbf{I}_d &= \text{diag}(\mathbf{S} - \Gamma\Gamma') \end{aligned}$$

Operando con la traza en la tercera

$$\begin{aligned} \mathbf{d}\sigma^2 &= \text{Tr}(\mathbf{S}) - \text{Tr}(\Gamma\Gamma') = \text{Tr}(\mathbf{S}) - \text{Tr}(\Gamma'\Gamma) = \text{Tr}(\mathbf{S}) - \text{Tr}(\Lambda_m) + \mathbf{m}\sigma^2 \\ (\mathbf{d}-\mathbf{m})\sigma^2 &= \sum_{i=1}^{\mathbf{d}} \lambda_i - \sum_{i=1}^{\mathbf{m}} \lambda_i = \sum_{i=\mathbf{m}+1}^{\mathbf{d}} \lambda_i \end{aligned}$$

Luego tomo

$$\widehat{\sigma}^2 = \sum_{i=\mathbf{m}+1}^{\mathbf{d}} \lambda_i / (\mathbf{d}-\mathbf{m})$$

En la segunda  $\Gamma'\Gamma = \Lambda_m - \sigma^2\mathbf{I}_m$  (**diagonal**)

$$(\Lambda_m - \sigma^2\mathbf{I}_m)^{-1/2} \Gamma'\Gamma (\Lambda_m - \sigma^2\mathbf{I}_m)^{-1/2} = \mathbf{I}_m$$

Luego si llamo  $\mathbf{T}_m$  a los primeros autovectores normalizados de la primera, resulta  $\mathbf{T}_m' \mathbf{T}_m = \mathbf{I}_m$

con

$$\Gamma (\Lambda_m - \sigma^2\mathbf{I}_m)^{-1/2} = \mathbf{T}_m \Rightarrow \Gamma = \mathbf{T}_m (\Lambda_m - \sigma^2\mathbf{I}_m)^{1/2}$$

Luego en definitiva queda

$$\hat{\Gamma} = \mathbf{T}_m (\Lambda_m - \hat{\sigma}^2 \mathbf{I}_m)^{1/2}$$

### Observación

Puedo también **no** aplicar las ecuaciones de máxima verosimilitud, y resolverlo con componentes principales así:

Sea  $\Sigma$  con autovalores  $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m, \underbrace{\sigma^2, \dots, \sigma^2}_{\text{(los últimos } \mathbf{d-m} \text{ autovectores iguales)}}$

Según componentes principales descompongo

$$\mathbf{x} = \mathbf{T}_m \mathbf{T}_m' \mathbf{x} + (\mathbf{I}_d - \mathbf{T}_m \mathbf{T}_m') \mathbf{x} \quad (\text{en partes ortogonales})$$

Luego  $\Sigma = \mathbf{T}_m \mathbf{T}_m' \Sigma \mathbf{T}_m \mathbf{T}_m' + (\mathbf{I}_d - \mathbf{T}_m \mathbf{T}_m') \Sigma (\mathbf{I}_d - \mathbf{T}_m \mathbf{T}_m')$

Pero  $\mathbf{I}_d = \mathbf{T}_m \mathbf{T}_m' + \mathbf{T}_{d-m} \mathbf{T}_{d-m}' \Rightarrow \Sigma = \underbrace{\mathbf{T}_m \mathbf{T}_m' \Sigma \mathbf{T}_m \mathbf{T}_m'}_{\Lambda_m} + \mathbf{T}_{d-m} \underbrace{\mathbf{T}_{d-m}' \Sigma \mathbf{T}_{d-m} \mathbf{T}_{d-m}'}_{\Lambda_{d-m} = \sigma^2 \mathbf{I}_{d-m}}$

$$\Sigma = \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \sigma^2 \mathbf{T}_{d-m} \mathbf{T}_{d-m}'$$

$$\Sigma = \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \sigma^2 (\mathbf{I}_d - \mathbf{T}_m \mathbf{T}_m')$$

$$\Sigma = \mathbf{T}_m (\Lambda_m - \sigma^2 \mathbf{I}_m) \mathbf{T}_m' + \sigma^2 \mathbf{I}_d$$

Que conduce a la misma solución que antes.

Este método se puede aplicar, en forma aproximada, cuando los últimos **d-m** autovalores son parecidos, o en el caso que sean diferentes, aporten poco a la varianza total.

### (c) Método de los factores principales (Thomson 1934)

Supongo el modelo  $\Sigma = \Gamma \Gamma' + \Psi$  y conozco **S**

Este método es similar al de máxima verosimilitud, salvo que aquí se impone la condición  $\Gamma' \Gamma = \mathbf{D}$  sea diagonal (unicidad por rotaciones).

Maximizando la verosimilitud respecto de  $\Gamma$  y  $\Psi$  (sujeto a la restricción  $\Gamma' \Gamma$  diagonal) resultan las ecuaciones (Thomson)

$$\begin{aligned} (\mathbf{S} - \Psi) \Gamma &= \Gamma \mathbf{D} \\ \Gamma' \Gamma &= \mathbf{D} \text{ (diagonal)} \\ \Psi &= \text{diag}(\mathbf{S} - \Gamma \Gamma') \end{aligned}$$

En la primera, **D** tiene los primeros **m** autovalores de  $\mathbf{S} - \Psi$ , y  $\Gamma$  tendrá los primeros **m** autovectores normalizados según la segunda ecuación.

#### Algoritmo iterativo

1°) Parto de una estimación inicial  $\hat{\Psi}_0$

2°) Como  $\Sigma - \Psi = \Gamma \Gamma'$  debe ser 1/2 definida positiva de rango **m**, descompongo

$$\mathbf{S} - \hat{\Psi}_0 = \mathbf{T}_m \Lambda_m \mathbf{T}_m' = \mathbf{T}_m \Lambda_m^{1/2} \Lambda_m^{1/2} \mathbf{T}_m' = \mathbf{T}_m \Lambda_m^{1/2} (\mathbf{T}_m \Lambda_m^{1/2})'$$

Virgilio L. Foglia (2003)

De aquí sale una primera estimación  $\hat{\Gamma}^0 = \mathbf{T}_m \Lambda_m^{1/2}$

3°) Re-estimamos  $\hat{\Psi}^1 = \mathbf{diag}(\mathbf{S} - \hat{\Gamma}^0 \hat{\Gamma}^{0'})$

y luego con 2°) obtenemos  $\hat{\Gamma}^1$

y con 3°)  $\hat{\Psi}^2$ , y así sucesivamente.

### Criterio de convergencia

En cada paso se evalúan las communalidades de las variables que están en  $\mathbf{diag}(\hat{\Gamma}^i \hat{\Gamma}^{i'})$ , y se calcula el máximo cambio, o sea

$$\Delta_i = \max \left\{ \text{valores absolutos de los elementos de } \mathbf{diag}(\hat{\Gamma}^i \hat{\Gamma}^{i'} - \hat{\Gamma}^{i-1} \hat{\Gamma}^{i-1'}) \right\}$$

Se itera hasta que  $\Delta_i$  sea menor que cierta cota.

**Problema:** En principio  $\Sigma - \Psi = \Gamma \Gamma'$  debe tener **d-m** autovalores nulos, sin embargo en la práctica,  $\mathbf{S} - \hat{\Psi}_0$ , puede no ser 1/2 definida positiva, y tener algunos autovalores negativos. Es crucial entonces la elección de **m**.

Como la communalidad total estimada inicialmente es  $\text{Tr}(\mathbf{S} - \hat{\Psi}_0)$ , la sugerencia entonces es elegir **m** de manera de acercarse a esta communalidad total, sin exceder el número de autovalores positivos.

### (e) Máxima Verosimilitud cuando $\Gamma$ es ident. fijando ciertos $\gamma_{ij} = 0$

Aquí no se requerirá que  $\Sigma_f = \Phi$  sea diagonal, pero sí que los elementos de la diagonal sean

1.

Asumimos que los "ceros" en  $\Gamma$  logran identificación.

Definimos una matriz  $\mathbf{H} \in \mathbf{R}^{d \times m}$  que tiene "ceros" en donde  $\Gamma$  no los tiene.

Se demuestra que las ecuaciones son:

$$\begin{aligned} (\Phi^{-1} + \Gamma' \Psi^{-1} \Gamma) \mathbf{H}' \Psi &= \Gamma' \Psi^{-1} \Sigma - \Gamma' - \Gamma' \Psi^{-1} \Phi \Gamma' \\ \mathbf{H}' \Gamma &= \mathbf{0} \\ \Psi &= \mathbf{diag}(\Sigma - \Gamma \Phi \Gamma') \quad (\text{diagonal}) \end{aligned}$$

### (f) Estimación cuando los factores no son aleatorios

Tengo  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  con  $\mathbf{x}_a = \Gamma \mathbf{f}_a + \mathbf{u} + \mu$  donde los  $\mathbf{f}_a$  son vectores fijos.

Esto es similar a cuadrados mínimos, salvo que las variables "independientes" son desconocidas, son parámetros a estimar.

$\Gamma$ ,  $\mu$  y  $\Psi$  son **parámetros estructurales** ya que afectan a **todas** las  $\mathbf{x}_a$ .

$\mathbf{f}_a$  son **parámetros específicos**, ya que afectan **solo** a la  $\mathbf{x}_a$  correspondiente.

Si suponemos que  $\mathbf{u} \sim \mathbf{N}_d(\mathbf{0}; \Psi)$

$$L = (2\pi)^{-dn/2} |\Psi|^{-n/2} \exp \left[ -\frac{1}{2} \sum_a (\mathbf{x}_a - \Gamma \mathbf{f}_a - \mu)' \Psi^{-1} (\mathbf{x}_a - \Gamma \mathbf{f}_a - \mu) \right]$$

Pero esta función no tiene máximo  $\rightarrow \nexists$  estimador de máxima verosimilitud.

Pero si bien no podemos usar MV a la distribución de las  $\mathbf{x}_a$ , sí podemos usar MV en la distribución de  $\hat{\Sigma}$ , que es una Whishart no central.

Con ciertas restricciones a  $\Gamma$  (para solucionar el problema de las rotaciones),  $\Gamma$  y  $\Psi$  resultan identificadas.

Se puede demostrar que las estimaciones por MV, suponiendo **factores aleatorios**, o **factores fijos con la Wishart no central**, son asintóticamente equivalentes.

Esto quiere decir que

$$\begin{aligned} \sqrt{n} (\hat{\Gamma}_w - \hat{\Gamma}_f) &\rightarrow \mathbf{0} \\ \sqrt{n} (\hat{\Psi}_w - \hat{\Psi}_f) &\rightarrow \mathbf{0} \end{aligned}$$

## Unidades de medida

Queremos analizar qué pasa, cuando cambiamos las unidades de medida de las  $\mathbf{x}$ .

Supongamos que en lugar de  $\mathbf{x}$  usamos  $\mathbf{x}^* = \mathbf{J}\mathbf{x}$ , donde  $\mathbf{J}$  es una matriz diagonal,  $\mathbf{J} \neq c\mathbf{I}_d$ .

Como  $\mathbf{x} = \Gamma \mathbf{f} + \varepsilon$

luego  $\mathbf{J}\mathbf{x} = \mathbf{J}\Gamma \mathbf{f} + \mathbf{J}\varepsilon$

$$\begin{aligned} \mathbf{x}^* = \Gamma^* \mathbf{f} + \varepsilon^* &\Rightarrow \Sigma^* = \Gamma^* \Gamma^{*\prime} + \Psi^* \\ \text{con } \Gamma^* &= \mathbf{J}\Gamma \text{ y } \Psi^* = \mathbf{J}\Psi\mathbf{J} \end{aligned}$$

o sea si usamos  $\mathbf{x}^* = \mathbf{J}\mathbf{x}$ , deberíamos obtener  $\Gamma^* = \mathbf{J}\Gamma$  y  $\Psi^* = \mathbf{J}\Psi\mathbf{J}$

Sin embargo esto no siempre es así, ya que esto depende del tipo de restricción que usemos para lograr unicidad, y del método de estimación.

## Dependencia respecto del tipo de restricción

(a) Forma triangular de  $\Gamma$  ( $\Gamma^*$ )

Si  $\Gamma$  es triangular, también lo será  $\mathbf{J}\Gamma$ ,  $\rightarrow \Gamma^* = \mathbf{J}\Gamma$

(b) Condición general de triangularidad

Si  $\mathbf{B}'\Gamma$  es triangular, en general  $\mathbf{B}'\mathbf{J}\Gamma$  no lo será  $\rightarrow \Gamma^* \neq \mathbf{J}\Gamma$

(c) Diagonalidad de  $\Gamma'\Gamma$  ( $\Gamma^{*\prime}\Gamma^*$ )

Si  $\Gamma'\Gamma$  es diagonal, en general  $(\mathbf{J}\Gamma)'\mathbf{J}\Gamma = \Gamma'\mathbf{J}^2\Gamma$  no lo será  $\rightarrow \Gamma^* \neq \mathbf{J}\Gamma$

(d) Diagonalidad de  $\Gamma'\Psi^{-1}\Gamma$  ( $\Gamma^{*\prime}\Psi^{*-1}\Gamma^*$ )

Si  $\Gamma'\Psi^{-1}\Gamma$  es diagonal, también lo será  $(\mathbf{J}\Gamma)'\mathbf{J}\Psi\mathbf{J}^{-1}(\mathbf{J}\Gamma) = \Gamma'\Psi^{-1}\Gamma \rightarrow \Gamma^* = \mathbf{J}\Gamma$

## Dependencia respecto del método de estimación

(a) Máxima verosimilitud

Si uso  $\mathbf{x}^* = \mathbf{J}\mathbf{x} \rightarrow \mathbf{S}^* = \mathbf{J}\mathbf{S}\mathbf{J}$  luego las ecuaciones a resolver en  $\Gamma^*$  y  $\Psi^*$  son

$$\begin{aligned} \mathbf{J}\mathbf{S}\mathbf{J}\Psi^{*-1}\Gamma^* &= \Gamma^*(\mathbf{I}_m + \mathbf{D}^*) \\ \Gamma^{*\prime}\Psi^{*-1}\Gamma^* &= \mathbf{D}^* \text{ (diagonal)} \\ \Psi^* &= \text{diag}(\mathbf{J}\mathbf{S}\mathbf{J} - \Gamma^*\Gamma^{*\prime}) \end{aligned}$$

Reemplazando resulta que  $\Gamma^* = \mathbf{J}\Gamma$  y  $\Psi^* = \mathbf{J}\Psi\mathbf{J}$  son soluciones (cuando  $\Gamma$  y  $\Psi$  también lo son).

(b) Método de las componentes principales

Si tengo que  $\Psi = \sigma^2\mathbf{I}_d \rightarrow \Psi^* = \mathbf{J}\Psi\mathbf{J} = \sigma^2\mathbf{J}^2 \neq \sigma^{*2}\mathbf{I}_d$

Luego si el método es aplicable a  $\Psi$ , no lo será a  $\Psi^*$ , y recíprocamente.

(c) Método de los factores principales (Thomson)

Si uso  $\mathbf{x}^* = \mathbf{J}\mathbf{x} \rightarrow \mathbf{S}^* = \mathbf{J}\mathbf{S}\mathbf{J}$  luego las ecuaciones a resolver en  $\Gamma^*$  y  $\Psi^*$  son

$$\begin{aligned} (\mathbf{J}\mathbf{S}\mathbf{J} - \Psi^*)\Gamma^* &= \Gamma^* \mathbf{D}^* \\ \Gamma^{*\prime} \Gamma^* &= \mathbf{D}^* \text{ (diagonal)} \\ \Psi^* &= \text{diag}(\mathbf{J}\mathbf{S}\mathbf{J} - \Gamma^* \Gamma^{*\prime}) \end{aligned}$$

Pero si  $\Gamma' \Gamma$  es diagonal, en general  $(\mathbf{J}\Gamma)' \mathbf{J}\Gamma = \Gamma' \mathbf{J}^2 \Gamma$  no lo será  $\rightarrow \Gamma^* \neq \mathbf{J}\Gamma$

**Una inquietud:** Si bien  $\mathbf{J}\Gamma$  no será solución, acaso una rotación  $\Gamma^\Delta = \mathbf{J}\Gamma\mathbf{L}$ , no lo será?

Si lo fuese  $\Gamma^\Delta \Gamma^{\Delta\prime} = \mathbf{J}\Gamma\mathbf{L}\mathbf{L}'\Gamma' \mathbf{J} = \mathbf{J}\Gamma\Gamma' \mathbf{J}$

y como debe ser  $\Sigma^\Delta = \Gamma^\Delta \Gamma^{\Delta\prime} + \Psi^\Delta \rightarrow \mathbf{J}\Sigma\mathbf{J} = \mathbf{J}\Gamma\Gamma' \mathbf{J} + \Psi^\Delta \rightarrow \Psi^\Delta = \mathbf{J}\Psi\mathbf{J}$

luego  $(\mathbf{J}\mathbf{S}\mathbf{J} - \mathbf{J}\Psi\mathbf{J})\Gamma^\Delta = \Gamma^\Delta \mathbf{D}^\Delta$

$$\mathbf{J}(\mathbf{S} - \Psi)\mathbf{J}\Gamma^\Delta = \Gamma^\Delta \mathbf{D}^\Delta$$

$$(\mathbf{S} - \Psi)\mathbf{J}^2 \mathbf{J}^{-1} \Gamma^\Delta = \mathbf{J}^{-1} \Gamma^\Delta \mathbf{D}^\Delta$$

Entonces  $\mathbf{D}^\Delta$  tendrá los  $m$  mayores autovalores de  $(\mathbf{S} - \Psi)\mathbf{J}^2$ . Pero estos serán en general diferentes de los de  $(\mathbf{S} - \Psi)$ . Luego los autovalores generarían un espacio factorial diferente.

O sea, cambiando las unidades de medida, cambiará el espacio de los factores en el método de Thomson.

(e) Máxima verosimilitud cuando  $\Gamma$  es identificada fijando ciertos  $\gamma_{ij} = 0$

Es claro que si  $\Gamma$  tiene "ceros" en ciertas posiciones, también los tendrá  $\Gamma^* = \mathbf{J}\Gamma$ .

También  $\mathbf{H}$  los tendrá en los mismos lugares que  $\mathbf{H}^*$ .

Reemplazando en las ecuaciones  $\mathbf{S}$  por  $\mathbf{J}\mathbf{S}\mathbf{J}$  se demuestra que

$$\Gamma^* = \mathbf{J}\Gamma, \Psi^* = \mathbf{J}\Psi\mathbf{J}, \Phi^* = \Phi, \mathbf{H}^* = \mathbf{J}^{-1}\mathbf{H}.$$

(f) Estimación cuando los factores no son aleatorios

Se demuestra que la verosimilitud en  $\mathbf{S}$ ,  $\Gamma$ , y  $\Psi$  es la misma que en  $\mathbf{J}\mathbf{S}\mathbf{J}$ ,  $\mathbf{J}\Gamma$  y  $\mathbf{J}\Psi\mathbf{J}$ .

Luego el máximo para  $\mathbf{S}^* = \mathbf{J}\mathbf{S}\mathbf{J}$ , estará en  $\Gamma^* = \mathbf{J}\Gamma$  y  $\Psi^* = \mathbf{J}\Psi\mathbf{J}$ , cuando el máximo para  $\mathbf{S}$  esté en  $\Gamma$  y  $\Psi$ .

## 6) PROBLEMAS DE INFERENCIA

### Test de ajuste del modelo

Para evaluar la bondad del ajuste, testaremos si la matriz de covarianza  $\Sigma$  tiene la forma  $\Gamma\Gamma' + \Psi$ , contra la hipótesis que es cualquier matriz definida positiva.

$$\boxed{\mathbf{H}_0: \Sigma = \Gamma\Gamma' + \Psi \quad \mathbf{K}: \Sigma > \mathbf{0}}$$

Como  $\ln L(\mathbf{S}, \hat{\Sigma}) = c - \frac{n}{2} (\ln |\hat{\Sigma}| - \text{Tr}(\hat{\Sigma}^{-1} \mathbf{S}))$

$$C_v = -2 (\ln L(\mathbf{S}, \hat{\Gamma}\hat{\Gamma}' + \hat{\Psi}) - \ln L(\mathbf{S}, \mathbf{S}))$$

$$C_v = -2 \left[ -\frac{n}{2} (\ln |\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi}| + \text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1} \mathbf{S})) - \ln |\mathbf{S}| - \underbrace{\text{Tr}(\mathbf{S}^{-1} \mathbf{S})}_{\mathbf{d}} \right]$$

**Lemita**  $\boxed{\text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1}\mathbf{S}) = \mathbf{d}}$

$$\begin{aligned} \text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1}\mathbf{S}) &= \text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1}\mathbf{S}) - \text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1}(\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})) + \mathbf{d} \\ &= \text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi})) + \mathbf{d} \end{aligned}$$

usando la relación:  $(\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1} = \hat{\Psi}^{-1} - \hat{\Psi}^{-1}\hat{\Gamma}(\mathbf{I}_m + \mathbf{D})^{-1}\hat{\Gamma}'\hat{\Psi}^{-1}$

$$\begin{aligned} &= \text{Tr}((\hat{\Psi}^{-1} - \hat{\Psi}^{-1}\hat{\Gamma}(\mathbf{I}_m + \mathbf{D})^{-1}\hat{\Gamma}'\hat{\Psi}^{-1})(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi})) + \mathbf{d} \\ &= \text{Tr}(\hat{\Psi}^{-1}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi})) - \text{Tr}(\hat{\Psi}^{-1}\hat{\Gamma}(\mathbf{I}_m + \mathbf{D})^{-1}\hat{\Gamma}'\hat{\Psi}^{-1}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi})) + \mathbf{d} \end{aligned}$$

Pero de la primera ecuacion de máxima verosimilitud

$$\begin{aligned} (\mathbf{S} - \hat{\Psi})\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\Gamma}\mathbf{D} \\ (\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi} + \hat{\Gamma}\hat{\Gamma}')\hat{\Psi}^{-1}\hat{\Gamma} &= \hat{\Gamma}\mathbf{D} \\ (\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi})\hat{\Psi}^{-1}\hat{\Gamma} + \underbrace{\hat{\Gamma}\hat{\Gamma}'\hat{\Psi}^{-1}\hat{\Gamma}}_{\mathbf{D}} &= \hat{\Gamma}\mathbf{D} \end{aligned}$$

$$(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi})\hat{\Psi}^{-1}\hat{\Gamma} = \mathbf{0} \rightarrow \text{la segunda traza es cero}$$

y además como de la tercera ecuación de máxima verosimilitud

$$\hat{\Psi} = \text{diag}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}') \rightarrow \text{diag}(\mathbf{S} - \hat{\Gamma}\hat{\Gamma}' - \hat{\Psi}) = \mathbf{0} \rightarrow \text{la primera traza es cero (pues } \hat{\Psi}^{-1} \text{ es diagonal)}$$

**Observación:** Jöreskog (1967) propuso, para estimar los parámetros en un modelo factorial, en lugar de maximizar la verosimilitud, minimizar

$$C_v = -2 \left[ -\frac{n}{2} (\ln |\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi}| + \text{Tr}((\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi})^{-1}\mathbf{S}) - \ln |\mathbf{S}| - \text{Tr}(\mathbf{S}^{-1}\mathbf{S})) \right]$$

o sea

$$C_v = -2 \left[ -\frac{n}{2} (\ln |\hat{\Sigma}| + \text{Tr}(\hat{\Sigma}^{-1}\mathbf{S}) - \ln |\mathbf{S}| - \mathbf{d}) \right]$$

o sea

$$\boxed{\text{mín } \Delta(\mathbf{S}, \hat{\Sigma}) = \ln |\hat{\Sigma}| + \text{Tr}(\hat{\Sigma}^{-1}\mathbf{S}) - \ln |\mathbf{S}| - \mathbf{d}}$$

Con esto se obtienen idénticos resultados, pero mejora mucho el algoritmo de estimación.

Continuando con la demostración queda

$$\begin{aligned} C_v &= -2 \left[ -\frac{n}{2} (\ln |\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi}| - \ln |\mathbf{S}|) \right] = n (\ln |\hat{\Gamma}\hat{\Gamma}' + \hat{\Psi}| - \ln |\mathbf{S}|) \\ &= n (\ln |\hat{\Sigma}| - \ln |\mathbf{S}|) \\ &= n \ln \left| \frac{\hat{\Sigma}}{|\mathbf{S}|} \right| \end{aligned}$$

Asintoticamente se tiene

$$C_v = n \ln \left| \frac{\hat{\Sigma}}{|S|} \right| \sim \chi_C^2 \text{ con } C = \frac{1}{2} \left[ (\mathbf{d} - \mathbf{m})^2 - (\mathbf{d} + \mathbf{m}) \right]$$

(Barlett propone reemplazar  $n$ , por  $n - \frac{2d+11}{6} - \frac{2m}{3}$ )

## Determinación del número de factores

En muchos casos el investigador no conoce el número de factores.

El problema entonces, es determinar el número de factores, tal que el correspondiente modelo ajuste bien a los datos.

Un método podría ser la secuencia de tests:

1°)  $H: \mathbf{m} = \mathbf{m}_0$        $K: \mathbf{m} > \mathbf{m}_0$

2°)  $H: \mathbf{m} = \mathbf{m}_0 + 1$      $K: \mathbf{m} > \mathbf{m}_0 + 1$  etc.

Pero no se conocería la significación global de esta secuencia de tests.

Otro sería usando el test de razón de verosimilitud:

1°)  $H: \mathbf{m} = \mathbf{m}_0$        $K: \mathbf{m} = \mathbf{m}_0 + 1$

2°)  $H: \mathbf{m} = \mathbf{m}_0 + 1$      $K: \mathbf{m} = \mathbf{m}_0 + 2$  etc.

aparecerían  $C_v(\mathbf{m}_0) - C_v(\mathbf{m}_0 + 1)$ ,  $C_v(\mathbf{m}_0 + 1) - C_v(\mathbf{m}_0 + 2)$  pero no se conoce la distribución asintótica de estos estadísticos.

## 7) ESTIMACION DEL SCORE DE LOS FACTORES

### Factor aleatorio (modelo condicional)

Consideremos el modelo  $\mathbf{x}^* = \Gamma \mathbf{f} + \varepsilon$  ( centrado con  $\mathbf{x}^* = \mathbf{x} - \bar{\mathbf{x}}$  )

y que con una muestra  $\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_n^*$  estimamos  $\hat{\Gamma}$  y  $\hat{\Psi}$ .

Ahora queremos estimar los factores para cada observación, o sea  $\hat{\mathbf{f}}_i$  para cada  $\mathbf{x}_i^*$ .

Notar que suponiendo normalidad, la distribución condicional

$$\mathbf{x}^* / \mathbf{f} \sim N_d(\Gamma \mathbf{f}; \Psi)$$

Luego condicional a los factores,  $\mathbf{x}^* = \Gamma \mathbf{f} + \varepsilon$  es un modelo lineal.

Aplicando cuadrados mínimos generalizados, y reemplazando  $\Gamma$  y  $\Psi$  por sus estimaciones

$$\hat{\mathbf{f}} = (\hat{\Gamma}' \hat{\Psi}^{-1} \hat{\Gamma})^{-1} \hat{\Gamma}' \hat{\Psi}^{-1} \mathbf{x}^*$$

Luego

$$\hat{\mathbf{f}}_i = (\hat{\Gamma}' \hat{\Psi}^{-1} \hat{\Gamma})^{-1} \hat{\Gamma}' \hat{\Psi}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$$

y si usamos maxima verosimilitud para estimar  $\Gamma$  y  $\Psi$ , resultará  $\hat{\Gamma}' \hat{\Psi}^{-1} \hat{\Gamma} = \mathbf{D}$  luego

$$\hat{\mathbf{f}}_i = \mathbf{D}^{-1} \hat{\Gamma}' \hat{\Psi}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$$

### Factor aleatorio (modelo no condicional)

Ahora supondremos el modelo usual , no condicional, con  $\mathbf{f}$  aleatorio.



Virgilio L. Foglia (2003)

$$\mathbf{x}^* = \Gamma \mathbf{f} + \varepsilon$$

y buscamos una transformación lineal  $\hat{\mathbf{f}} = \mathbf{A}\mathbf{x}^*$  que minimize el ECM o sea

$$\begin{aligned} & \mathbf{E} \left\| \hat{\mathbf{f}} - \mathbf{f} \right\|^2 \text{ mínimo} \\ & \mathbf{E} \left\| \mathbf{A}\mathbf{x}^* - \mathbf{f} \right\|^2 = \mathbf{E} \left\| \mathbf{A}(\Gamma \mathbf{f} + \varepsilon) - \mathbf{f} \right\|^2 \\ & = \mathbf{E} \left\| (\mathbf{A}\Gamma - \mathbf{I}_m)\mathbf{f} + \mathbf{A}\varepsilon \right\|^2 \\ & = \mathbf{E} \left[ \mathbf{f}' (\mathbf{A}\Gamma - \mathbf{I}_m)' (\mathbf{A}\Gamma - \mathbf{I}_m) \mathbf{f} \right] + 2 \underbrace{\mathbf{E} \left[ \varepsilon' \mathbf{A}' (\mathbf{A}\Gamma - \mathbf{I}_m) \mathbf{f} \right]} + \mathbf{E} (\varepsilon' \mathbf{A}' \mathbf{A} \varepsilon) \end{aligned}$$

pues  $\varepsilon$  y  $\mathbf{f}$  son no correlacionadas de media cero

Ahora usando que  $\mathbf{E}(\mathbf{z}' \mathbf{H} \mathbf{z}) = \text{Tr}(\mathbf{H}\Sigma) + \mu' \mathbf{H} \mu$

$$\begin{aligned} & = \text{Tr} \left[ (\mathbf{A}\Gamma - \mathbf{I}_m)' (\mathbf{A}\Gamma - \mathbf{I}_m) \underbrace{\Sigma_{\mathbf{f}}}_{\mathbf{I}_m} \right] + \text{Tr}(\mathbf{A}' \mathbf{A} \underbrace{\Sigma_{\varepsilon}}_{\Psi}) \\ & = \text{Tr}(\Gamma' \mathbf{A}' \mathbf{A} \Gamma + \mathbf{I}_m - \Gamma' \mathbf{A}' - \mathbf{A} \Gamma) + \text{Tr}(\mathbf{A}' \mathbf{A} \Psi) \\ & = \text{Tr}(\mathbf{A}' \mathbf{A} \underbrace{\Gamma \Gamma'}_{\Sigma - \Psi}) + \mathbf{m} - 2\text{Tr}(\mathbf{A} \Gamma) + \text{Tr}(\mathbf{A}' \mathbf{A} \Psi) \\ & = \text{Tr}(\mathbf{A}' \mathbf{A} \Sigma) - \text{Tr}(\mathbf{A}' \mathbf{A} \Psi) + \mathbf{m} - 2\text{Tr}(\mathbf{A} \Gamma) + \text{Tr}(\mathbf{A}' \mathbf{A} \Psi) \\ & = \text{Tr}(\mathbf{A}' \mathbf{A} \Sigma) - 2\text{Tr}(\mathbf{A} \Gamma) + \mathbf{m} \end{aligned}$$

derivando respecto de  $\mathbf{A}$

$$2\mathbf{A}\Sigma - 2\Gamma' = \mathbf{0} \rightarrow \mathbf{A} = \Gamma' \Sigma^{-1} \text{ (Thomson 1951)}$$

Luego queda  $\hat{\mathbf{f}} = \Gamma' \Sigma^{-1} \mathbf{x}^*$  y reemplazando por las estimaciones

$$\boxed{\hat{\mathbf{f}}_i = \hat{\Gamma}' \hat{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})}$$

Usando la relación  $\Gamma' \Sigma^{-1} = (\mathbf{I}_m + \Gamma' \Psi^{-1} \Gamma)^{-1} \Gamma' \Psi^{-1}$

**D**

Queda  $\hat{\mathbf{f}}_i = (\mathbf{I}_m + \mathbf{D})^{-1} \hat{\Gamma}' \hat{\Psi}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$  mientras que para el modelo condicional

quedaba  $\hat{\mathbf{f}}_i = \mathbf{D}^{-1} \hat{\Gamma}' \hat{\Psi}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$

luego los dos estimadores difieren en un múltiplo escalar.

Además este estimador del modelo no condicional, es un "ridge estimador" en el modelo condicional, y por lo tanto es sesgado pero tiene menor error cuadrático medio.

## 8) ROTACION DE LOS FACTORES

### Rotación ortogonal

Virgilio L. Foglia (2003)

Aquí suponemos que  $\Phi = \mathbf{I}_m$ , o sea, los factores son ortogonales entre sí, y que ya hemos estimado el modelo, obteniendo  $\hat{\Gamma}$  y  $\hat{\Psi}$ .

Buscamos una matriz ortogonal  $\mathbf{L}$ , tal que  $\hat{\Gamma}^* = \hat{\Gamma}\mathbf{L}$ , permita una mejor interpretación de los factores.

Como  $\hat{\Gamma}^* \hat{\Gamma}^{*'} = \hat{\Gamma} \hat{\Gamma}'$ , las communalidades de las variables se mantendrán iguales luego de la rotación,

Sin embargo como  $\hat{\Gamma}^* \hat{\Gamma}^{*'} = \mathbf{L}' \hat{\Gamma}' \hat{\Gamma} \mathbf{L} \neq \hat{\Gamma}' \hat{\Gamma}$ , las varianzas totales explicadas por cada factor en general serán diferentes.

### Rotación Quartimax

El objetivo aquí es lograr una matriz  $\hat{\Gamma}^*$  que tenga el aspecto (por ejemplo para  $m=4$  y  $d=8$ )

$$\hat{\Gamma}^* = \begin{bmatrix} x & x & \cdot & \cdot \\ x & \cdot & x & \cdot \\ x & x & \cdot & \cdot \\ x & \cdot & \cdot & x \\ x & \cdot & x & \cdot \\ x & \cdot & x & \cdot \\ x & \cdot & \cdot & x \\ x & x & \cdot & \cdot \end{bmatrix} \quad \text{con: "x"} \rightarrow \text{carga alta} \quad \text{y} \quad \text{"\cdot"} \rightarrow \text{carga pequeña}$$

O sea queremos que todas las variables tengan una carga mas bien alta en un mismo factor, y que además tengan a lo sumo carga alta en un factor, y pequeña en los otros. Una estructura de este tipo supone que las variables se pueden interpretar en función de un factor general, y algunos factores específicos.

Si llamamos

$$Q_i = \text{var}(\gamma_{i1}^2, \gamma_{i2}^2, \dots, \gamma_{im}^2) = \frac{1}{m} \sum_{j=1}^m (\gamma_{ij}^2 - \gamma_i^2)^2 = \frac{1}{m} \sum_{j=1}^m \gamma_{ij}^4 - \frac{1}{m^2} \underbrace{\left( \sum_{j=1}^m \gamma_{ij}^2 \right)^2}_{h_i^2}$$

$$= \frac{1}{m} \sum_{j=1}^m \gamma_{ij}^4 - h_i^2/m^2$$

El criterio es maximizar  $Q = Q_1 + Q_2 + \dots + Q_d = \frac{1}{m} \sum_{i=1}^d \sum_{j=1}^m \gamma_{ij}^4 - \frac{1}{m^2} \sum_{i=1}^d h_i^2$

sujeto a la restricción que las  $h_i^2$  se mantengan fijas.

Luego queda

$$\boxed{\text{máx} \sum_{i=1}^d \sum_{j=1}^m \gamma_{ij}^4 \quad \text{con la restricción } h_i^2 \text{ fijas para todo } i}$$

### Rotación varimáx

El objetivo aquí es lograr una matriz  $\hat{\Gamma}^*$  que tenga el aspecto (por ejemplo para  $m=4$  y  $d=8$ )

Virgilio L. Foglia (2003)

$$\hat{\Gamma}^* = \begin{bmatrix} \cdot & x & \cdot & \cdot \\ x & \cdot & \cdot & \cdot \\ \cdot & \cdot & x & \cdot \\ \cdot & \cdot & \cdot & x \\ \cdot & \cdot & x & \cdot \\ \cdot & \cdot & x & \cdot \\ \cdot & \cdot & \cdot & x \\ \cdot & x & \cdot & \cdot \end{bmatrix} \quad \text{con: "x"} \rightarrow \text{carga alta} \quad \text{y " \cdot " } \rightarrow \text{carga pequeña}$$

O sea queremos que cada variable tengan una carga alta en solo un factor, y que en los restantes tenga una carga muy pequeña.

Si llamamos

$$V_j = \text{var}(\gamma_{1j}^2, \gamma_{2j}^2, \dots, \gamma_{dj}^2) = \frac{1}{d} \sum_{i=1}^d (\gamma_{ij}^2 - \gamma_{\cdot j}^2)^2 = \frac{1}{d} \sum_{i=1}^d \gamma_{ij}^4 - \frac{1}{d^2} \left( \sum_{i=1}^d \gamma_{ij}^2 \right)^2$$

El criterio es maximizar  $V = V_1 + V_2 + \dots + V_m$  sujeto a la restricción que las  $h_i^2$  se mantengan fijas.

$$\text{Pero } V = \frac{1}{d} \sum_{j=1}^m \sum_{i=1}^d \gamma_{ij}^4 - \frac{1}{d^2} \sum_{j=1}^m \left( \sum_{i=1}^d \gamma_{ij}^2 \right)^2$$

Tomando dV queda el criterio

$$\boxed{\text{máx} \sum_{j=1}^m \sum_{i=1}^d \gamma_{ij}^4 - \frac{1}{d} \sum_{j=1}^m \left( \sum_{i=1}^d \gamma_{ij}^2 \right)^2 \quad \text{con la restricción } h_i^2 \text{ fijas para todo } i}$$

### Rotación Equimáx, y Biquartimáx

Vimos que la rotación Quartimáx maximiza la varianza total de las cargas al cuadrado de las filas de  $\hat{\Gamma}$ , mientras que la varimáx lo hace con las columnas. Podemos optar por una solución intermedia entre estas, tomando un promedio ponderado

$$Z = \alpha Q + \beta dV \quad \text{con } \alpha \text{ y } \beta \text{ dados}$$

Reemplazando, operando, y llamando  $\theta = \beta/(\alpha + \beta)$ , queda el criterio

$$\boxed{\text{máx} \sum_{j=1}^m \sum_{i=1}^d \gamma_{ij}^4 - \theta \sum_{j=1}^m \left( \sum_{i=1}^d \gamma_{ij}^2 \right)^2 \quad \text{con la restricción } h_i^2 \text{ fijas para todo } i}$$

Luego para

$$\theta = m/2 \rightarrow \text{Rotación Equimáx}$$

$$\theta = 1/2 \rightarrow \text{Rotación Biquartimáx}$$

## Ejemplo: Análisis Factorial vs. Componentes Principales

Ya fué analizado el método de las componentes principales para hacer la descomposición factorial. Esencialmente consistía en el de máxima verosimilitud cuando se puede suponer que  $\Psi = \sigma^2 \mathbf{I}_d$ . Este sí es un método del análisis factorial.

Sin embargo es muy común en los textos, y en los soft's (caso SPSS) proponer el empleo de la descomposición de la matriz de covarianza en sus componentes principales, como método aproximado. La idea es la siguiente (operando directamente sobre  $\mathbf{S}$ , y suponiendo que en  $\Lambda_m$  están los primeros  $m$  autovalores de  $\mathbf{S}$ )

$$\begin{aligned} \mathbf{S} = \mathbf{T}\Lambda\mathbf{T}' &= \begin{bmatrix} \mathbf{T}_m & \mathbf{T}_{d-m} \end{bmatrix} \begin{bmatrix} \Lambda_m & \mathbf{0} \\ \mathbf{0} & \Lambda_{d-m} \end{bmatrix} \begin{bmatrix} \mathbf{T}_m' \\ \mathbf{T}_{d-m}' \end{bmatrix} = \mathbf{T}_m \Lambda_m \mathbf{T}_m' + \mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}' \\ &= \mathbf{T}_m \Lambda_m^{1/2} (\mathbf{T}_m \Lambda_m^{1/2})' + \mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}' \\ &\simeq \mathbf{T}_m \Lambda_m^{1/2} (\mathbf{T}_m \Lambda_m^{1/2})' + \mathbf{diag}(\mathbf{T}_{d-m} \Lambda_{d-m} \mathbf{T}_{d-m}') \\ &= \Gamma_{CP} \Gamma_{CP}' + \Psi_{CP} \end{aligned}$$

Notar que se reemplaza la última matriz, que en realidad NO es diagonal, por otra que solo contiene sus elementos diagonales.

Compararemos este método con el de máxima verosimilitud en el ejemplo ya citado, sobre las calificaciones de fin de curso, de 200 estudiantes, en Matemática (M), Física (F), Química (Q), Inglés (I), Historia (H) y Alemán (A).

La matriz de correlación muestral entre las calificaciones es

$$\mathbf{S} = \begin{bmatrix} 1 & 0.62 & 0.54 & 0.32 & 0.284 & 0.37 \\ 0.62 & 1 & 0.51 & 0.38 & 0.351 & 0.43 \\ 0.54 & 0.51 & 1 & 0.36 & 0.336 & 0.405 \\ 0.32 & 0.38 & 0.36 & 1 & 0.686 & 0.73 \\ 0.284 & 0.351 & 0.336 & 0.686 & 1 & 0.735 \\ 0.37 & 0.43 & 0.405 & 0.73 & 0.735 & 1 \end{bmatrix}$$

Luego extrayendo 2 factores, resultan según los dos métodos las estimaciones

$$\hat{\Gamma}_{MV} = \begin{bmatrix} 0.572 & 0.594 \\ 0.609 & 0.458 \\ 0.559 & 0.371 \\ 0.793 & -0.224 \\ 0.786 & -0.278 \\ 0.862 & -0.206 \end{bmatrix} \quad \hat{\Psi}_{MV} = \begin{bmatrix} 0.320 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.420 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.550 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.320 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.305 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.215 \end{bmatrix}$$

$$\hat{\Gamma}_{CP} = \begin{bmatrix} 0.675 & 0.557 \\ 0.718 & 0.447 \\ 0.683 & 0.418 \\ 0.793 & -0.410 \\ 0.774 & -0.461 \\ 0.838 & -0.359 \end{bmatrix} \quad \hat{\Psi}_{CP} = \begin{bmatrix} 0.234 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.284 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.359 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.203 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.188 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.169 \end{bmatrix}$$

Se observan ciertas diferencias, la más evidente es que las varianzas residuales con el método de componentes principales son menores. Esto es debido a que las componentes principales maximizan la varianza explicada por las componentes (factores aquí), minimizando

por tanto los residuos. En cambio en análisis factorial, la intención es explicar lo mejor posible las correlaciones entre las variables, a través de los factores comunes.

Las matrices de covarianza explicadas por los factores son:

$$\hat{\Gamma}_{MV} \hat{\Gamma}_{MV}' = \begin{bmatrix} .680 & .6204 & .54012 & .32054 & .28446 & .3707 \\ .6204 & .580 & .51035 & .38035 & .35135 & .43061 \\ .54012 & .51035 & .450 & .36018 & .33624 & .40543 \\ .32054 & .38035 & .36018 & .680 & .68557 & .72971 \\ .28446 & .35135 & .33624 & .68557 & .695 & .7348 \\ .3707 & .43061 & .40543 & .72971 & .7348 & .785 \end{bmatrix}$$

$$\hat{\Gamma}_{CP} \hat{\Gamma}_{CP}' = \begin{bmatrix} .766 & .73363 & .69385 & .30691 & .26567 & .36569 \\ .73363 & .716 & .67724 & .3861 & .34967 & .44121 \\ .69385 & .67724 & .641 & .37024 & .33594 & .42229 \\ .30691 & .3861 & .37024 & .797 & .80279 & .81172 \\ .26567 & .34967 & .33594 & .80279 & .812 & .81411 \\ .36569 & .44121 & .42229 & .81172 & .81411 & .831 \end{bmatrix}$$

En las diagonales están las comunales de cada variable. Si calculamos la comunalidad total explicada por ambos factores resulta

$$\mathbf{h}_{MV}^2 = 0.680 + 0.580 + 0.450 + 0.680 + 0.695 + 0.785 = 3.87$$

$$\mathbf{h}_{CP}^2 = 0.766 + 0.716 + 0.641 + 0.797 + 0.812 + 0.831 = 4.56$$

como la varianza total de las variables observadas es 6 (matriz de correlación) luego los porcentajes de la varianza total explicada por los factores son:

$$\%var_{MV} = 100 * 3.87/6 = 64.5 \%$$

$$\%var_{CP} = 100 * 4.56/6 = 76 \%$$

Por iguales motivos que antes las componentes principales explican una proporción mayor de la varianza total.

A continuación figuran las matrices de covarianzas explicadas por los modelos

$$\hat{\Sigma}_{MV} = \hat{\Gamma}_{MV} \hat{\Gamma}'_{MV} + \hat{\Psi}_{MV} = \begin{bmatrix} 1 & .6204 & .54012 & .32054 & .28446 & .3707 \\ .6204 & 1 & .51035 & .38035 & .35135 & .43061 \\ .54012 & .51035 & 1 & .36018 & .33624 & .40543 \\ .32054 & .38035 & .36018 & 1 & .68557 & .72971 \\ .28446 & .35135 & .33624 & .68557 & 1 & .7348 \\ .3707 & .43061 & .40543 & .72971 & .7348 & 1 \end{bmatrix}$$

$$\hat{\Sigma}_{CP} = \hat{\Gamma}_{CP} \hat{\Gamma}'_{CP} + \hat{\Psi}_{CP} = \begin{bmatrix} 1 & .73363 & .69385 & .30691 & .26567 & .36569 \\ .73363 & 1 & .67724 & .3861 & .34967 & .44121 \\ .69385 & .67724 & 1 & .37024 & .33594 & .42229 \\ .30691 & .3861 & .37024 & 1 & .80279 & .81172 \\ .26567 & .34967 & .33594 & .80279 & 1 & .81411 \\ .36569 & .44121 & .42229 & .81172 & .81411 & 1 \end{bmatrix}$$

Y para comparar el ajuste de los modelos es útil la matriz residual, entre  $\mathbf{S}$  y la estimada por el modelo  $\hat{\Sigma}$ , o sea  $\mathbf{R} = \mathbf{S} - \hat{\Sigma}$

$$\hat{\mathbf{R}}_{MV} = \mathbf{S} - \hat{\Sigma}_{MV} = \begin{bmatrix} 0 & -.0004 & -.00012 & -.00054 & -.00046 & -.0007 \\ -.0004 & 0 & -.00035 & -.00035 & -.00035 & -.00061 \\ -.00012 & -.00035 & 0 & -.00018 & -.00024 & -.00043 \\ -.00054 & -.00035 & -.00018 & 0 & .00043 & .00029 \\ -.00046 & -.00035 & -.00024 & .00043 & 0 & .0002 \\ -.0007 & -.00061 & -.00043 & .00029 & .0002 & 0 \end{bmatrix}$$

$$\hat{\mathbf{R}}_{CP} = \mathbf{S} - \hat{\Sigma}_{CP} = \begin{bmatrix} 0 & -.11363 & -.15385 & .01309 & .01833 & .00431 \\ -.11363 & 0 & -.16724 & -.0061 & .00133 & -.01121 \\ -.15385 & -.16724 & 0 & -.01024 & .00006 & -.01729 \\ .01309 & -.0061 & -.01024 & 0 & -.11679 & -.08172 \\ .01833 & .00133 & .00006 & -.11679 & 0 & -.07911 \\ .00431 & -.01121 & -.01729 & -.08172 & -.07911 & 0 \end{bmatrix}$$

Notar aquí una gran diferencia. Los elementos fuera de la diagonal principal miden la diferencia entre las correlaciones observadas y las explicadas por los modelos. Claramente es muy superior el análisis factorial (aquí MV), para explicar las correlaciones, y si este es el objetivo del problema, no importa tanto el porcentaje de la varianza total que explican los factores.

## 9) BIBLIOGRAFIA

- ♦ **Statistical inference in factor analysis**  
T.W.Anderson & Herman Rubin

Virgilio L. Foglia (2003)

(Third Berkeley Symposium)

◆ **Multivariate Observations**

G.A.F. Seber

(John Wiley 1984)

◆ **Applied Multivariate Statistical Techniques**

Subhash Sharma

(John Wiley 1996)

◆ **An Introduction to Latent Variable Models**

B.S. Everitt

(Chapman and Hall 1984)

◆ **Basic Principles of Structural Equation Modeling**

Ralph O. Mueller

(Springer-Verlag 1996)