

El análisis de factores o análisis factorial tiene como objetivo explicar un conjunto de variables observadas mediante un pequeño número de variables latentes, o no observadas (llamados factores), con la mínima pérdida de información

## Ejemplo 1:

- Supongamos que se toman 20 medidas físicas del cuerpo de una persona: longitud del tronco y de las extremidades, anchura de hombros, peso, etc.
- Las mediciones no son independientes entre sí, conocidas algunas de ellas se pueden prever las restantes con poco error.
- Esto se puede explicar por el hecho de que las dimensiones del cuerpo humano dependen de ciertos factores, que si son conocidos, se podrían prever las dimensiones con poco error.

## Ejemplo 2:

- Se está interesado en estudiar el desarrollo humano en los países del mundo y se disponen de muchas variables económicas, sociales y demográficas.
- En general, todas las variables son dependientes entre si, que están relacionadas con el desarrollo.
- Podemos preguntarnos si el desarrollo de un país depende de un pequeño número de factores tales que, conocidos sus valores, podríamos prever el conjunto de las variables de cada país

## Ejemplo 3:

- Se mide la capacidad mental de un individuo para procesar información y resolver problemas, mediante distintas pruebas.
- Existen algunos factores no directamente observables, que expliquen los resultados observados?
- El conjunto de estos factores será lo que llamamos *inteligencia*.
- Es importante conocer cuantas dimensiones distintas tiene este concepto y como caracterizarlas y medirlas.

## Ejemplo 4:

- Se desea medir la capacidad de *abstracción*, *analítica* y *memoria* de un grupo de estudiantes universitarios.
- Para lo cual, cada alumno fue evaluado en las siguientes 10 *areas de conocimientos*:

Álgebra	Contabilidad financiera
Cálculo	Análisis de costos
Estadística	Comunicación comercial
Derecho mercantil	Actuariales
Derecho laboral	Econometría

- Se observaron las 10 calificaciones de cada alumno y entre estas notas, o al menos entre algunas de ellas, se observan correlaciones elevadas que, en cierta medida, provienen de aptitudes globales del alumno que no se observan directamente de las áreas evaluadas.

## Ejemplo 4:

- Un análisis factorial podría permitir que la información relativa a estas variables se resumiese en tres únicos factores, sin pérdida excesiva de información
- Cada uno de estos tres factores se interpretaría como:

$F_1$  — Factor de CAPACIDAD DE ABSTRACCIÓN

$F_2$  — Factor de MEMORIA

$F_3$  — Factor de CAPACIDAD ANALÍTICA

**El primer modelo de factores fue propuesto por Karl Pearson y Charles Spearman, en su interés por comprender las dimensiones de la inteligencia humana.**

# Análisis de Factores: origen

- Karl Pearson y Charles Spearman derivaron el análisis de factores bajo el siguiente argumento. Supongamos que las variables se pueden agrupar de acuerdo a sus correlaciones, es decir, supongamos que todas las variables dentro de un grupo particular están altamente correlacionadas entre ellas, pero tienen una correlación pequeña con variables en un grupo distinto.
- Entonces es posible que cada grupo de variables represente una *única dimensión subyacente o factor*, que es responsable de las correlaciones observadas en cada grupo.
- Este tipo de estructura es lo que el análisis de factores busca identificar.

# Análisis de Factores: origen relacionado con el área de psicología

## Primeros trabajos de referencia del Análisis de factores

- Holzinger, K. J. and Swineford, F. (1939). A study in Factor Analysis: The stability of a Bi-Factor solution, University of Chicago: Supplementary Educational Monographs, Number 48.
- Spearman, C. (1904). General intelligence objectively determined and measured. American Journal of Psychology, 15, 201-293.
- Thurstone, L. L. (1935). Vectors of the mind. Chicago: University of Chicago Press.
- Thurstone, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

# Objetivos de la aplicación del Análisis de factores

- **Seleccionar un subconjunto de variables** entre un gran número de opciones, con base en la selección de las más altas correlaciones entre las variables originales y los factores.
- **Generar un grupo de factores no correlacionados** como una forma de enfrentar la presencia de colinealidades en procedimientos de regresión múltiple.
- **Identificar grupos de individuos**
- **Identificar valores extremos o outliers**
- **Predecir valores faltantes**



# Análisis de Factores y Análisis de componentes principales (PCA)

- El análisis de factores se puede considerar una extensión de componentes principales.
- Ambas técnicas intentan reducir la dimensionalidad de los datos mediante un número pequeño de términos, de tal forma que éstos capturen de manera aproximada la variabilidad de los datos originales.
- Sin embargo, la manera en la que AF construye estos términos es mas elaborada. Al igual que PCA, Análisis de factores intenta explicar las correlaciones entre variable, pero considerando que existe una fuente de variación común a las variables y otra fuente de variación específica de cada variable.
- PCA es una herramienta descriptiva, mientras que AF presupone un modelo estadístico formal de generación de datos.

## 1 Análisis de factores exploratorio

- Busca descubrir la estructura *subyacente* de un gran número de variables mediante un conjunto de factores no observables
- No se conoce a priori el número de factores. El investigador supone a priori que existen indicadores que estarán asociados a algún factor.
- No existe una teoría a priori para identificar estas asociaciones sino que se interpretan las cargas de los factores para descubrir esas estructuras

## 2 Análisis de factores confirmatorio

- Se asume que existen ciertos factores que explican la estructura de los datos de acuerdo a una teoría pre-establecida.
- Las variables indicadoras se seleccionan con base en la teoría preestablecida
- Se busca determinar o confirmar si los factores y sus cargas corresponden con lo que establece la teoría
- El nombre y número de los factores puede definirse a priori.

# Áreas de aplicación de análisis de factores

- Es una de las técnicas multivariadas más utilizadas y con más literatura disponible.
- Su aplicación es muy amplia, cubriendo áreas como mercadotecnia, psicología, educación, procesos de producción, salud, recursos humanos, desarrollos de nuevos productos, etc.

## **Algunas referencias de la aplicación de análisis de factores:**

- Pett, Marjorie A., Nancy R. Lackey, and John J. Sullivan (2003). Making sense of factor analysis: The use of factor analysis for instrument development in health care research. Thousand Oaks, CA: Sage Publications
- Widaman, K. F. (1993). Common factor analysis versus principal components analysis: Differential bias in representing model parameters?" Multivariate Behavioral Research 28: 263-311. Cited with regard to preference for PFA over PCA in confirmatory factor analysis in SEM.

# Variabilidad en Mediciones

- Al igual que PCA, los modelos de análisis de factores se construyen basados en la variabilidad observada en las mediciones de variables.
- Sin embargo en AF se asume que la variabilidad de cada variable  $X_i$  se descompone en dos términos: la varianza común y la varianza única (específica).
  - *La varianza común* es la parte de la variación de la variable que es compartida con las otras variables.
  - *La varianza única (específica)* es la parte de la variación de la variable que es propia de esa variable.
- Generalmente se refiere al AF como el análisis que busca un nuevo conjunto de variables, menor en número que las variables originales, que exprese lo que es común a esas variables e identificando también su variabilidad específica.

# Etapas del análisis de factores

- 1 **Recolección de datos y generación de matriz de covarianzas/correlaciones.** No es necesario disponer de los datos originales, la matriz de covarianzas/correlaciones es suficiente
- 2 **Obtención de factores.** Se obtienen las combinaciones lineales que forman los factores ortogonales, sin embargo los factores pueden ser no interpretables.
- 3 **Realizar una rotación de los factores para una interpretación adecuada.** Se busca una transformación que *cambie* los coeficientes de manera que facilite la interpretación de los factores. Existen varios métodos para encontrar la rotación más conveniente.
- 4 **Generación de escalas o puntajes de factores (factor scores)** para utilizarlos en análisis adicionales.

# Modelo de factores ortogonales

- Consideremos un vector aleatorio observable  $\mathbf{X}$ , con  $p$  componentes, que tiene vector de medias  $\mu$  y matriz de covarianzas  $\Sigma$ .
- El modelo de factores establece que las variables observables de  $\mathbf{X}$  son generadas mediante una combinación lineal de  $m$  variables aleatorias no observables

$$F_1, F_2, \dots, F_m (m < p)$$

llamadas *factores comunes* y  $p$  fuentes de variación adicionales

$$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$$

llamados *errores* o también *factores específicos*

# Modelo de factores ortogonales

Algebraicamente el modelo de análisis de factores se establece de la siguiente forma:

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \varepsilon_2$$

$$X_3 - \mu_3 = l_{31}F_1 + l_{32}F_2 + \cdots + l_{3m}F_m + \varepsilon_3$$

$$\vdots$$

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \varepsilon_p$$

O bien en notación matricial

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$$

# Modelo de factores ortogonales

- $L$  representa la matriz de cargas de los factores y cada una de sus entradas,  $l_{ij}$ , representa la carga de la  $i$ -ésima variable en el  $j$ -ésimo factor
- El  $i$ -ésimo factor específico  $\varepsilon_i$  está asociado solo con la  $i$ -ésima respuesta  $X_i$
- Los  $p$  elementos del vector de desviaciones  $(\mathbf{X} - \mu)$  están expresados en términos de  $m + p$  variables aleatorias

$$F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$$

las cuales son *no observables*

- Está característica distingue al modelo de análisis de factores del modelo de regresión lineal múltiple:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

donde  $\mathbf{X}$  es un vector de variables observables.



# Modelo de factores ortogonales

Además se consideran los siguientes supuestos acerca de los vectores aleatorios  $\mathbf{F}$  y  $\varepsilon$ :

$$E(\mathbf{F}) = \mathbf{0} \quad \text{Cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \mathbf{I}$$

$$E(\varepsilon) = \mathbf{0} \quad \text{Cov}(\varepsilon) = E(\varepsilon\varepsilon') = \Psi$$

donde

$$\Psi = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

$\mathbf{F}$  y  $\varepsilon$  son independientes, es decir,

$$\text{Cov}(\varepsilon, \mathbf{F}) = E(\varepsilon\mathbf{F}') = E(\varepsilon)E(\mathbf{F}') = \mathbf{0}.$$

Así que la ecuación matricial del modelo, *junto* con estos supuestos constituyen el modelo de *factores ortogonales*.

# Modelo de factores ortogonales

De los supuestos anteriores, podemos expresar la matriz de covarianzas de  $\mathbf{X}$  como sigue:

$$\begin{aligned}\Sigma &= \text{Cov}(\mathbf{X}) \\ &= E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)'] \\ &= E[(\mathbf{L}\mathbf{F} + \varepsilon)(\mathbf{L}\mathbf{F} + \varepsilon)'] \\ &= E[\mathbf{L}\mathbf{F}\mathbf{F}'\mathbf{L}' + \mathbf{L}\mathbf{F}\varepsilon' + \varepsilon\mathbf{F}'\mathbf{L}' + \varepsilon\varepsilon'] \\ &= \mathbf{L}\mathbf{L}' + \Psi\end{aligned}$$

Con los mismos supuestos, se puede también concluir que:

$$\begin{aligned}\text{Cov}(\mathbf{X}, \mathbf{F}) &= E[(\mathbf{X} - \mu)(\mathbf{F} - 0)'] = E[(\mathbf{X} - \mu)\mathbf{F}'] \\ &= E[(\mathbf{L}\mathbf{F} + \varepsilon)\mathbf{F}'] = E[\mathbf{L}\mathbf{F}\mathbf{F}' + \varepsilon\mathbf{F}'] \\ &= E[\mathbf{L}\mathbf{F}\mathbf{F}'] \\ &= \mathbf{L}\end{aligned}$$

# Modelo de factores ortogonales

De la relación

$$\Sigma = LL' + \Psi$$

Las varianzas y covarianzas entre las  $X_i$  se pueden expresar por las cargas y las varianzas de los factores específicos (errores) de la siguiente forma:

$$Var(X_i) = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i$$

$$Cov(X_i, X_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \cdots + l_{im}l_{km}$$

Se puede ver que cada covarianza entre una variable original y un factor común es la carga entre ellos

$$Cov(X_i, F_j) = l_{ij}$$

# Modelo de factores ortogonales

La varianza de la  $i$ -ésima variable se puede separar en dos partes:

- 1 La contribución de los factores comunes, expresada por la suma de cuadrados de las cargas en esa variable, que se denomina **comunalidad**.
- 2 La contribución del factor específico, llamada **varianza específica**.

$$\begin{aligned} \text{Var}(X_i) &= \sigma_{ii} \\ &= l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i \\ &= h_i^2 + \psi_i \end{aligned}$$

$$\text{Comunalidad : } h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2$$

$$\text{Varianza específica : } \psi_i$$

# Modelo de factores ortogonales

- El modelo de factores asume que las  $p + p(p-1)/2 = p(p+1)/2$  varianzas y covarianzas para  $\mathbf{X}$  se pueden reproducir a partir de las  $pm$  cargas  $l_{ij}$  y de las  $p$  varianzas específicas  $\psi_i$ .
- Cuando el número de factores es igual al número de variables, es decir,  $m = p$ , cualquier matriz de covarianzas  $\Sigma$  se puede reproducir exactamente como  $\mathbf{LL}'$ , de tal modo que  $\Psi$  sería una matriz de ceros, lo cual resulta poco práctico, debido a que la idea es reducir la dimensionalidad original de los datos
- Cuando  $m = 1$ , en muchas situaciones las matrices de covarianzas  $\Sigma$  no pueden ser factorizadas como  $\mathbf{LL}' + \Psi$ , y en caso de que si se pueda obtener esta factorización, las soluciones no serán consistentes con la interpretación de los resultados.

# Rotación de los factores y cargas

- Cuando  $m > 1$ , siempre existirá alguna ambigüedad asociada con el modelo de factores. Si  $\mathbf{T}$  es cualquier matriz ortogonal  $m \times m$ , es decir,  $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$ , entonces el modelo se puede escribir como

$$\begin{aligned}\mathbf{X} - \boldsymbol{\mu} &= \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} \\ &= \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{F} + \boldsymbol{\varepsilon} \\ &= \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\varepsilon}\end{aligned}$$

- Es decir, el modelo puede ser re-expresado a través de una rotación de las cargas y factores, estas rotaciones están dadas por:

$$\begin{aligned}\mathbf{L}^* &= \mathbf{L}\mathbf{T} \quad \text{y} \quad \mathbf{F}^* = \mathbf{T}'\mathbf{F} \\ E(\mathbf{F}^*) &= \mathbf{T}'E(\mathbf{F}) = \mathbf{0} \\ \text{Cov}(\mathbf{F}^*) &= \mathbf{T}'\text{Cov}(\mathbf{F})\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}_{m \times m}\end{aligned}$$

# Rotación de los factores y cargas

- Por tanto el nuevo factor  $F^*$  cumple los supuestos originales del modelo.
- Por tanto, la estructura de covarianza original  $\Sigma$  no se ve afectada por las rotaciones de las cargas y de los factores.
- Se busca una rotación adecuada que proporcione una fácil interpretación de los factores.
- Es imposible distinguir los pesos  $L$  de los  $L^*$  basado en las observaciones  $X$

$F$  y  $F^* = T'F$  tienen las mismas propiedades estadísticas

- A pesar que las cargas  $L^*$  son en general diferentes de las cargas  $L$ , ambas generan la misma matriz  $\Sigma$ , debido a que

$$\Sigma = LL' + \Psi = LTT'L' + \Psi = (L^*)(L^*)' + \Psi$$

# Rotación de los factores y cargas

- Esta ambigüedad justifica la rotación de factores, debido a que las matrices ortogonales representan rotaciones del sistema de coordenadas para  $\mathbf{X}$

## En resumen:

- Las matrices  $\mathbf{L}^*$  y  $\mathbf{L}$  dan la misma representación de  $\mathbf{\Sigma} = \mathbf{Cov}(\mathbf{X})$
- Las communalidades dadas por los elementos de la diagonal de  $\mathbf{LL}' = (\mathbf{L}^*)(\mathbf{L}^*)'$  no son afectadas por la elección de  $\mathbf{T}$ .
- Una vez obtenidas las cargas y los factores específicos es común obtener estimaciones para cada uno de los factores en cada uno de los  $n$  casos, conocidos como *factor scores*



# Conveniencia del análisis de factores

- Dado un conjunto de  $n$  observaciones vectoriales  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  en  $p$  variables
- ¿Cuándo es apropiado utilizar un modelo con un número pequeño de factores para representar los datos?
- Si los elementos fuera de la diagonal de  $\mathbf{S}$  son pequeños, o bien los elementos fuera de la diagonal de la matriz de correlaciones  $\mathbf{R}$  son esencialmente cero, las variables no se relacionan entre sí y un análisis de factores no será de gran utilidad.
- Bartlett propuso una prueba para determinar si las variables no están correlacionadas, denominada *prueba de esfericidad de Bartlett*
- El estadístico de prueba está dado por

$$- \left[ (n-1) - \frac{(2p+5)}{6} \right] \ln |\mathbf{R}| \sim \chi^2_{(p^2-p)/2}$$

# Prueba de Esfericidad de Bartlett

- El determinante de la matriz de correlaciones  $\mathbf{R}$  es igual al producto de sus  $p$  valores propios.
- Cuando  $\mathbf{R}$  es próxima a la matriz identidad su determinante es próximo a 1 y por lo tanto su logaritmo natural es cercano a cero.
- Cuando  $\mathbf{R}$  tiene valores propios próximos a cero, el determinante de  $\mathbf{R}$  es próximo a cero y su logaritmo natural es un valor negativo con magnitud grande.
- Valores grandes del estadístico rechazan la hipótesis nula de que la matriz de correlaciones es igual a la matriz identidad.
- Se busca rechazar la hipótesis de esfericidad para proseguir con un análisis de componentes principales o análisis de factores.

# Métodos de estimación de los parámetros del modelo de factores

Existen varios métodos para estimar los parámetros del modelo, los métodos mas utilizados son:

- ❶ Análisis de Componentes Principales(PCA, incluyendo el método del factor principal)
  - Método del factor principal: cuando se desea identificar variables latentes que contribuyan a la varianza común de las variables medidas, excluyendo la *varianza específica* (única).
- ❷ El Método de máxima verosimilitud, asumiendo normalidad de los datos

Es aconsejable aplicar los dos métodos de estimación, las soluciones deberían ser consistentes unas con otras. Ambos métodos requieren cálculos iterativos que deben ser hechos por computadora.

# Estimación por componentes principales

- Proporciona una solución única que permite reconstruir las varianzas y covarianzas de las variables originales a partir de  $p$  factores resultantes.
- Se obtienen tantos factores como variables originales.
- A través de diferentes criterios se elegirán los factores a considerar relevantes.
- Se utiliza generalmente cuando el objetivo es la reducción del número de componentes.
- Reproduce tanto la varianza común como la varianza específica de las variables en el estudio.

# Estimación por componentes principales

Se basa en la factorización de la matriz  $\Sigma$  mediante la descomposición espectral. Sea  $\Sigma$  con  $(\lambda_i, \mathbf{e}_i)$  sus pares de valores y vectores propios tal que

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

Entonces:

$$\begin{aligned} \Sigma &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \\ &= [ \sqrt{\lambda_1} \mathbf{e}_1 \mid \sqrt{\lambda_2} \mathbf{e}_2 \mid \dots \mid \sqrt{\lambda_p} \mathbf{e}_p ] \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p} \mathbf{e}_p' \end{bmatrix} = \mathbf{L} \mathbf{L}' \end{aligned}$$

Por lo tanto, ajustamos la estructura de covarianza para el modelo de análisis de factores, teniendo tantos factores como variables ( $m = p$ ) y varianzas específicas  $\psi_i = 0$  para toda  $i$ .

# Estimación por componentes principales

- La  $j$ -ésima columna de la matriz de cargas  $\mathbf{L}$  está dada por  $\sqrt{\lambda_j} \mathbf{e}_j$
- De esta forma se puede escribir

$$\underset{(p \times p)}{\mathbf{\Sigma}} = \underset{(p \times p)}{\mathbf{L}} \underset{(p \times p)}{\mathbf{L}'} + \underset{(p \times p)}{\mathbf{0}} = \mathbf{L} \mathbf{L}'$$

- Además del factor escalar  $\sqrt{\lambda_j}$ , las cargas factoriales en el  $j$ -ésimo factor son los coeficientes de la  $j$ -ésima componente principal de la población.

# Estimación por componentes principales

- Lo que se busca es representar la estructura de la matriz varianzas y covarianzas  $\Sigma$  en términos de pocos factores comunes.
- Una aproximación se da cuando los  $p - m$  valores propios son pequeños y entonces se elimina la contribución de

$$\lambda_{m+1} \mathbf{e}_{m+1} \mathbf{e}_{m+1}' + \lambda_{m+2} \mathbf{e}_{m+2} \mathbf{e}_{m+2}' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$$

de  $\Sigma$ .

- Eliminando esa contribución, se obtiene la aproximación a  $\Sigma$  como

$$\Sigma \doteq \left[ \sqrt{\lambda_1} \mathbf{e}_1 \mid \sqrt{\lambda_2} \mathbf{e}_2 \mid \cdots \mid \sqrt{\lambda_m} \mathbf{e}_m \right] \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} = \underset{(p \times m)}{\mathbf{L}} \underset{(m \times p)}{\mathbf{L}'}$$

# Estimación por componentes principales

- Esta representación supone que los factores específicos  $\varepsilon$  son de menor importancia y pueden ser eliminados de la factorización de  $\Sigma$ .
- Si se incluyen los factores específicos en el modelo, sus varianzas pueden ser tomadas como los elementos diagonales de  $\Sigma - LL'$  y la aproximación está dada por

$$\Sigma \doteq [ \sqrt{\lambda_1} \mathbf{e}_1 \mid \sqrt{\lambda_2} \mathbf{e}_2 \mid \cdots \mid \sqrt{\lambda_m} \mathbf{e}_m ] \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} + \begin{bmatrix} \hat{\psi}_1 & 0 & \cdots & 0 \\ 0 & \hat{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\psi}_p \end{bmatrix}$$

donde  $\hat{\psi}_i = \sigma_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2$  para  $i = 1, 2, \dots, p$



# Estimación por componentes principales

Para aplicar esta aproximación a una muestra de observaciones  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  es usual, primero, centrar las observaciones, restándoles la media muestral  $\bar{\mathbf{X}}$

$$\mathbf{x}_j - \bar{\mathbf{X}} = \begin{bmatrix} X_{j1} - \bar{X}_1 \\ X_{j2} - \bar{X}_2 \\ \vdots \\ X_{jp} - \bar{X}_p \end{bmatrix}, \quad j = 1, \dots, n$$

y en ocasiones se estandarizan

$$\mathbf{z}_j = \begin{bmatrix} \frac{X_{j1} - \bar{X}_1}{\sqrt{s_{11}}} \\ \frac{X_{j2} - \bar{X}_2}{\sqrt{s_{22}}} \\ \vdots \\ \frac{X_{jp} - \bar{X}_p}{\sqrt{s_{pp}}} \end{bmatrix}, \quad j = 1, \dots, n$$

Cuya matriz de covarianzas es la matriz de correlaciones muestrales  $\mathbf{R}$  de las observaciones  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$

# Estimación por componentes principales

- La estandarización evita el problema de tener una variable con gran varianza, que afecta la determinación de las cargas de los factores.
- Entonces esta representación, cuando se aplican a la matriz  $\mathbf{S}$  o a la matriz  $\mathbf{R}$ , se conoce como la solución del modelo de factores *por componentes principales*.
- El nombre proviene del hecho de que las cargas de los factores son los coeficientes *escalados* de unos cuantos componentes principales muestrales.

## Resultado:

- El análisis de factores por componentes principales de la matriz de covarianzas  $\mathbf{S}$  es especificado en términos de sus pares de valores y vectores propios  $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \dots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$ , donde

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$$

- Sea  $m < p$ , el número de factores comunes. Entonces la matriz de cargas estimada  $\tilde{\mathbf{L}}$  está dada por

$$\tilde{\mathbf{L}} = \left[ \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 \mid \dots \mid \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \right]$$

- Los vectores y valores propios obtenidos de  $\mathbf{S}$ , representan estimaciones de los verdaderos valores y vectores propios

**Resultado (Continuación)** Las varianzas específicas estimadas están dadas por los elementos de la diagonal de  $\mathbf{S} - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'$ , entonces:

$$\widetilde{\Psi} = \begin{bmatrix} \widetilde{\psi}_1 & 0 & \cdots & 0 \\ 0 & \widetilde{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{\psi}_p \end{bmatrix} \quad \text{con} \quad \widetilde{\psi}_i = s_{ii} - \sum_{j=1}^m \widetilde{l}_{ij}^2$$

Y las comunidades son estimadas por:

$$\widetilde{h}_i^2 = \widetilde{l}_{i1}^2 + \widetilde{l}_{i2}^2 + \cdots + \widetilde{l}_{im}^2$$

Cuando las observaciones son estandarizadas, el análisis de factores por componente principales se realiza a partir de  $\mathbf{R}$  en lugar de  $\mathbf{S}$

# Determinación del número de factores

- Para una solución por componentes principales, los estimadores de las cargas no cambian conforme aumenta el número de factores.
- Es decir, si se consideran por ejemplo 3 factores principales y se obtienen las cargas correspondientes, entonces si aumentamos el número de factores a 4, las cargas de los primeros 3 factores no cambian.

## **Cómo determinar el número de factores a utilizar?**

Si no se tienen consideraciones a priori, la elección de  $m$  se puede basar en los valores propios estimados, de la misma manera como se hace con componentes principales

# Criterios para determinar el número de factores

- Idealmente, las contribuciones de los primeros factores a las varianzas muestrales de las variables deben ser grandes
- De manera análoga a componentes principales, la contribución a la varianza muestral total,  $s_{11} + s_{22} + \dots + s_{pp} = \text{tr}(\mathbf{S})$ , del primer factor común está dada por:

$$\tilde{l}_{11}^2 + \tilde{l}_{21}^2 + \dots + \tilde{l}_{p1}^2 = (\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1)' (\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1) = \hat{\lambda}_1$$

ya que los vectores propios  $\hat{\mathbf{e}}_1$  tienen longitud unitaria. En general

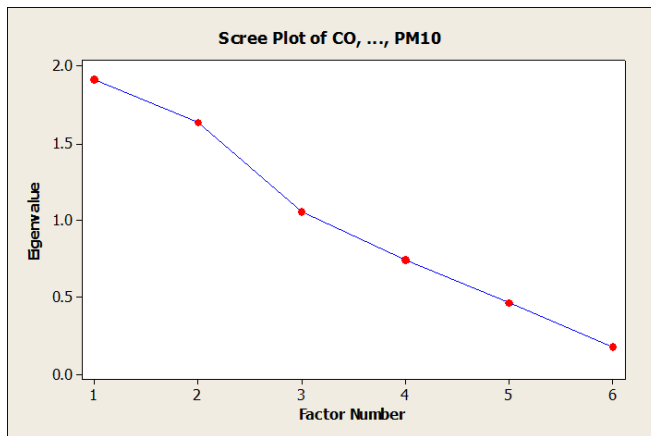
$$\left( \begin{array}{c} \text{Proporción de la} \\ \text{varianza muestral} \\ \text{total explicada} \\ \text{por el } j\text{-ésimo factor} \end{array} \right) = \left\{ \begin{array}{ll} \frac{\hat{\lambda}_j}{s_{11} + s_{22} + \dots + s_{pp}} & \text{para un análisis de factores de } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{para un análisis de factores de } \mathbf{R} \end{array} \right.$$

- Por lo general se elige el número de factores que explican de manera acumulada más del 70% de la varianza total

# Criterios para determinar el número de factores

- **Criterio de Kaiser.** Es uno de los criterios más utilizados por los paquetes estadísticos para determinar el número adecuado de factores. Se aplica cuando se trabaja con la matriz de correlaciones muestrales  **$R$** .
  - Consiste en elegir el número de factores como el número de valores propios de  **$R$**  mayores a 1.0
  - Esta regla busca que cualquier factor retenido contenga al menos la varianza de una de las variables utilizadas en el análisis.
  - Cuando el número de variables es pequeño, el número de valores propios  $> 1$  será pequeño, por tanto esta regla puede detectar menos factores que los que existen realmente.
  - Cuando el número de variables es grande, esta regla puede detectar más factores que los que realmente existen.

# Criterios para determinar el número de factores





# Otros criterios para determinar el número de factores

**Parallel analysis (PA)**, conocido también como *Humphrey-Ilggen parallel analysis*.

- Algunos estudios lo refieren como el mejor método (Velicer, Eaton, and Fava, 2000: 67; Lance, Butts, and Michels, 2006).
- Mediante este método se realiza un análisis de factores sobre una matriz de datos generada aleatoriamente y cuyas variables no están correlacionadas. Se considera el mismo número de casos y el mismo número de variables que tienen los datos reales.
- Se comparan gráficamente las dos líneas del scree plot, una para el análisis de los datos reales y otra para el análisis de los datos aleatorios.
- El número de factores elegido es aquel donde las dos líneas se *intersectan*.

# Otros criterios para determinar el número de factores

**Criterio de comparación de matrices de correlación.** Se obtiene la diferencia entre la matriz de correlaciones real y la matriz reproducida mediante  $m$  factores, si el número de factores  $m$  es adecuado, la diferencia debe ser cercana a cero.

- **Correlaciones reproducidas.** Es la matriz de correlaciones de las variables originales que resultaría suponiendo que es correcto el número de los factores retenidos.
- **Matriz de correlaciones residuales.** Es la matriz de las diferencias entre la matriz de correlaciones reproducida y la matriz de correlaciones reales. Si el número de factores  $m$  es adecuado, se esperaría que la matriz de correlaciones residuales tenga todas sus entradas cercanas a cero.

El mejor enfoque es retener pocos en lugar de muchos factores, suponiendo que proporcionan una interpretación satisfactoria de los datos y dan un ajuste satisfactorio a  **$S$**  o  **$R$** .

# Ejemplo: Estudio de preferencias de consumidores

- En un estudio de preferencias de consumidores, se pidió a una muestra aleatoria de clientes que evaluaran varios atributos de un nuevo producto. Las respuestas dadas por cada cliente en una escala de 7 puntos se tabularon y se construyó la matriz de correlación de los atributos. Los atributos (variables) medidos y la matriz de correlación son los siguientes:

**Atributos(Variables)**

Sabor

Barato, económico

Condimento

Adecuado como tentempié

Proporciona mucha energía

$$R = \begin{bmatrix} 1.00 & .02 & .96 & .42 & .01 \\ .02 & 1.00 & .13 & .71 & .85 \\ .96 & .13 & 1.00 & .50 & .11 \\ .42 & .71 & .50 & 1.00 & .79 \\ .01 & .85 & .11 & .79 & 1.00 \end{bmatrix}$$

- Construye el modelo de factores adecuado mediante componentes principales

# Rotación de factores

- Una ventaja del análisis de factores es que podemos rotar los factores, para obtener una mejor interpretación de ellos.
- Todas las cargas factoriales iniciales se pueden rotar mediante una transformación ortogonal y reproducir la matriz de covarianzas (o de correlación).
- Una transformación ortogonal corresponde a una rotación de los ejes coordenados
- Por esta razón una transformación ortogonal de las cargas factoriales, así como la transformación ortogonal implicada de los factores es llamada rotación factorial

# Rotación de factores

- La rotación busca mejorar la interpretación de factores.
  - Al rotar los factores, la suma de valores propios no se altera, pero sí cambian los valores propios, el porcentaje de varianza explicada y los valores de cargas en cada factor.
- Existen diferentes métodos de rotación:
  - Rotaciones ortogonales. Estas rotaciones mantienen los factores ortogonales, son transformaciones rígidas
  - Rotaciones oblicuas. Estas rotaciones producen factores correlacionados, son transformaciones no rígidas.  
Frecuentemente las cargas rotadas no tienen una fácil interpretación

# Principales rotaciones ortogonales

- *Rotación Varimax*. Rotación ortogonal que maximiza la varianza de las cargas de factores al cuadrado sobre todas las variables. Para cualquier variable en particular este método tiende a generar cargas o altas o bajas en los factores. Este método, por lo tanto facilita la interpretación de los factores. Es la opción más utilizada.
- *Rotación Quartimax* es una rotación ortogonal que minimiza el número de factores que se requiere para explicar cada variable.
- *Rotación Equimax* es un método intermedio entre los criterios varimax y quartimax.

El procedimiento Varimax selecciona la transformación ortogonal  $\mathbf{T}$  que maximiza:

$$V = \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p \tilde{l}_{ij}^{*4} - \left( \sum_{i=1}^p \tilde{l}_{ij}^{*2} \right)^2 / p \right]$$

donde  $\tilde{l}_{ij}^* = \frac{\tilde{l}_{ij}}{\sqrt{h_i}}$  es el coeficiente de rotación escalado por la raíz cuadrada de las comunales

- El escalamiento en las cargas tiene el efecto de dar mayor peso a las variables con comunales pequeñas, para determinar una estructura sencilla.

# Rotación de factores

- Después de determinar la transformación  $\mathbf{T}$ , las cargas  $\tilde{l}_{ij}^*$  son multiplicadas por  $\hat{h}_j$  de tal forma que las communalidades originales se preserven.
- $V_j$  tiene una interpretación simple, es

$$V_j \propto \sum_{i=1}^m (\text{La varianza de los cuadrados de las cargas (rotadas) para el } j\text{-ésimo factor})$$

- Así, maximizar  $V_j$  equivale a *expandir*, tanto como sea posible, los cuadrados de las cargas sobre cada factor,
- Por lo tanto, esperamos encontrar grupos de coeficientes grandes y pequeños en cualquier columna de la matriz de cargas rotadas  $\hat{\mathbf{L}}^*$



# Método del factor principal

- El método del factor principal es una modificación al enfoque de estimación de las cargas mediante componentes principales.
- Aunque este procedimiento para obtener las cargas es muy similar al de componentes principales, la diferencia entre ellos radica en el enfoque y en la forma de obtener las estimaciones iniciales de los parámetros.
- La elección del número de factores es menos simple que en ACP, además el número de factores es menor al número de variables originales.
- Al igual que el método por componentes principales, el objetivo es identificar variables latentes que contribuyen a la varianza común de las variables medidas, excluyendo la varianza “específica” (única).
- El método se describirá en términos de la factorización de  $\mathbf{R}$ , aunque también es apropiado para  $\mathbf{S}$

# Método del factor principal

- Suponiendo que el modelo  $\mathbf{R} = \mathbf{LL}' + \psi$  está correctamente especificado por los  $m$  factores
- Los  $m$  factores explicarán los elementos fuera de la diagonal de la matriz  $\mathbf{R}$  así como la parte común (comunalidades) en los elementos diagonales de  $\mathbf{R}$  dados por

$$r_{ii} = 1 = h_i^2 + \psi_i$$

- Si la contribución del factor específico  $\psi_i$  es removida de la diagonal o equivalentemente el 1 es reemplazado por  $h_i^2$  entonces la matriz resultante es

$$\mathbf{R} - \psi = \mathbf{LL}'$$

- Suponiendo ahora que tenemos disponible la estimación inicial de las varianzas específicas, dadas por  $\psi_i^*$ . Entonces reemplazando el  $i$ -ésimo elemento de la diagonal de  $\mathbf{R}$  por  $h_i^{*2} = 1 - \psi_i^*$  se obtiene la matriz de correlaciones muestral *reducida*  $\mathbf{R}_r$  dada por

$$\mathbf{R}_r = \begin{bmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{12} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & h_p^{*2} \end{bmatrix}$$

# Método del factor principal

- Además de la variación muestral, todos los elementos de la matriz de correlaciones reducida  $\mathbf{R}_r$ , deberían ser explicados por los  $m$  factores comunes. En particular  $\mathbf{R}_r$  se puede factorizar aproximadamente como

$$\mathbf{R}_r \approx \mathbf{L}_r^* \mathbf{L}_r^{*'}$$

donde  $\mathbf{L}_r^* = \{l_{ij}^*\}$ , son las cargas estimadas.

- El método del factor principal* de análisis de factores usa las estimaciones de las cargas, dadas por

$$\mathbf{L}_r^* = \left[ \sqrt{\hat{\lambda}_1^*} \hat{\mathbf{e}}_1^* : \sqrt{\hat{\lambda}_2^*} \hat{\mathbf{e}}_2^* : \dots : \sqrt{\hat{\lambda}_m^*} \hat{\mathbf{e}}_m^* \right]$$

$$\psi_i^* = 1 - \sum_{j=1}^m l_{ij}^{*2}$$

donde  $(\hat{\lambda}_i^*, \hat{\mathbf{e}}_i^*)$ ,  $i = 1, 2, \dots, m$ , son los pares de valores y vectores propios (mas grandes) obtenidos de  $\mathbf{R}_r$ .

# Método del factor principal

- Es decir, con las cargas estimadas,  $L_r^*$ , podemos obtener nuevas estimaciones para las comunidades:

$$\tilde{h}_i^{*2} = \sum_{j=1}^m l_{ij}^{*2}$$

- La solución del factor principal se obtiene iterativamente, tomando las comunidades estimadas,  $\tilde{h}_i^{*2} = \sum_{j=1}^m l_{ij}^{*2}$ , como la estimación inicial para la siguiente etapa del proceso
- En la solución por componentes principales, la consideración sobre los valores propios  $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \dots, \hat{\lambda}_p^*$  nos ayudan a determinar el número adecuado de factores a retener.
- Sin embargo, en el método del factor principal algunos de los valores propios pueden ser negativos, debido a la forma como se estiman inicialmente las comunidades, las cuales dependen de la estimación inicial de las varianzas específicas

# Método del factor principal

- De manera ideal, se podría tomar el número de factores comunes igual al rango de la *matriz de correlaciones reducida*, sin embargo el rango no siempre está bien determinado a partir de  $\mathbf{R}_r$  y entonces se deben tomar en cuenta otros elementos.
- Existen varias formas de obtener una estimación inicial de las varianzas específicas  $\psi_i$ , la más popular cuando se trabaja con  $\mathbf{R}$ , es  $\psi_i^* = 1/r^{ii}$ , donde  $r^{ii}$  es el  $i$ -ésimo elemento diagonal de  $\mathbf{R}^{-1}$ .
- Entonces la estimación inicial de las comunalidades está dada por

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}}$$

- la cual representa el cuadrado del coeficiente de correlación múltiple entre  $X_i$  y las otras  $p - 1$  variables

# Método del factor principal

- La relación con el coeficiente de correlación múltiple significa que  $h_i^{*2}$  se puede calcular aún cuando  $\mathbf{R}$  no tenga rango completo.
- Si se considera la factorización de  $\mathbf{S}$ , la estimación inicial de la  $\psi_i^*$  utiliza  $s^{ii}$ , que son los elementos diagonales de  $\mathbf{S}^{-1}$ .
- Aunque el método por componentes principales para  $\mathbf{R}$  se puede considerar como un método del factor principal, donde las comunales iniciales son estimadas como la unidad o varianzas específicas igual a cero, los dos enfoques son distintos.
- Sin embargo en la práctica, los dos métodos producen frecuentemente cargas comparables si el número de variables es grande y el número factores comunes es pequeño.

# Comentarios generales sobre el método de estimación mediante componentes principales.

- Las estimaciones de las cargas obtenidas mediante el método de componentes principales no son invariantes. ante transformaciones lineales
- Es decir, no se obtiene necesariamente el mismo resultado trabajando con la matriz de covarianzas y con la matriz correlaciones.
- Una alternativa es estimar las cargas  $\mathbf{L}$  y la matriz de varianzas específicas  $\Psi$  mediante el método de máxima verosimilitud.
- A diferencia del método de componentes principales, los EMV de  $\mathbf{L}$  y  $\Psi$  son invariantes ante transformaciones lineales, aunque su obtención tiene un mayor costo computacional.

# Estimación por el método de máxima verosimilitud

- Asumiendo que los factores comunes  $\mathbf{F}$  y los factores específicos  $\varepsilon$  son normalmente distribuidos, se pueden obtener los estimadores de máxima verosimilitud de las cargas de los factores y de las varianzas específicas.
- Cuando  $\mathbf{F}_j$  y  $\varepsilon_j$  siguen una distribución normal multivariada, las observaciones  $\mathbf{X}_j = \boldsymbol{\mu} + \mathbf{L}\mathbf{F}_j + \varepsilon_j$  siguen una distribución normal  $p$ -dimensional con vector de medias  $\boldsymbol{\mu}$ , y matriz de covarianzas  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ .
- Asumiendo independencia entre las observaciones  $\mathbf{X}_j$ , la función de verosimilitud es:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right)tr[\boldsymbol{\Sigma}^{-1}(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})')]}$$



Que se puede expresar también como:

$$L(\mu, \Sigma) = (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right) \text{tr}[\Sigma^{-1}(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})')] } \\ \times (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right) (\bar{\mathbf{x}} - \mu)' \Sigma^{-1} (\bar{\mathbf{x}} - \mu)}$$

la cual depende de  $\mathbf{L}$  y  $\Psi$  debido a que  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ .

- Sin embargo, este modelo no está bien definido porque tenemos muchas opciones para elegir  $\mathbf{L}$  mediante rotaciones de la forma  $\mathbf{L}^* = \mathbf{L}\mathbf{T}$ . Se dice que el modelo de factores está indeterminado por rotaciones.
- Por tanto, es necesario imponer restricciones computacionales sobre  $\mathbf{L}$  que definan un modelo factorial único.

# Estimación por el método de máxima verosimilitud

- Existen varias formas de imponer restricciones sobre  $\mathbf{L}$  para identificar un modelo factorial único.
- Una normalización muy utilizada para que  $\mathbf{L}$  quede bien definida es imponer la condición de unicidad de la forma

$$\mathbf{L}'\Psi^{-1}\mathbf{L} = \mathbf{\Delta}$$

donde  $\mathbf{\Delta}$  es una matriz diagonal

- Con esta normalización los efectos de los factores sobre las variables, ponderados por las varianzas específicas, se hacen incorrelacionados y define una matriz de cargas de manera única
- Agregando esta restricción, se obtienen los estimadores de máxima verosimilitud,  $\hat{\mathbf{L}}$  y  $\hat{\Psi}$ , maximizando numéricamente a  $L(\mu, \Sigma)$ .

# Estimación por el método de máxima verosimilitud

- Es necesario utilizar métodos numéricos para obtener  $\hat{\mathbf{L}}$  y  $\hat{\Psi}$ , debido a que las ecuaciones de verosimilitud resultantes no tienen una forma analítica sencilla.
- Los paquetes estadísticos utilizan algoritmos iterativos para obtener las estimaciones de una manera rápida y eficiente.

En el siguiente resultado se resumen algunas características de los estimadores de máxima verosimilitud obtenidos

## Resultado

- Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  una muestra aleatoria de una población  $N_p(\mu, \Sigma)$ , donde  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$  es la matriz de covarianzas para el modelo de  $m$  factores comunes.
- Los estimadores de máxima verosimilitud  $\hat{\mathbf{L}}, \hat{\Psi}, \hat{\mu} = \bar{\mathbf{x}}$  maximizan la función de verosimilitud  $L(\mu, \Sigma)$  bajo la restricción de que  $\hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}}$  sea diagonal.

## Resultado (Continuación)

- Debido a la propiedad de invarianza de los EMV, funciones de  $\mathbf{L}$  y  $\boldsymbol{\Psi}$  son estimadas por las mismas funciones de  $\hat{\mathbf{L}}$  y  $\hat{\boldsymbol{\Psi}}$ .
- En particular, los estimadores de máxima verosimilitud de las comunales  $h_i^2$  están dados por

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2 \quad i = 1, 2, \dots, p$$

- Así, la proporción de la varianza total de la muestra explicada por el  $j$ -ésimo factor es

$$\frac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \cdots + \hat{l}_{pj}^2}{s_{11} + s_{22} + \cdots + s_{pp}}$$

## Resultado considerando la estandarización de las variables

- Si  $\mathbf{Z} = \mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ , entonces la matriz de covarianzas  $\boldsymbol{\rho}$  de  $\mathbf{Z}$  se representa por

$$\begin{aligned}\boldsymbol{\rho} &= \mathbf{V}^{-1/2} \boldsymbol{\Sigma} \mathbf{V}^{-1/2} = \mathbf{V}^{-1/2} (\mathbf{L} \mathbf{L}' + \boldsymbol{\Psi}) \mathbf{V}^{-1/2} \\ &= (\mathbf{V}^{-1/2} \mathbf{L} \mathbf{L}' + \mathbf{V}^{-1/2} \boldsymbol{\Psi}) \mathbf{V}^{-1/2} \\ &= (\mathbf{V}^{-1/2} \mathbf{L} \mathbf{L}' \mathbf{V}^{-1/2}) + \mathbf{V}^{-1/2} \boldsymbol{\Psi} \mathbf{V}^{-1/2} \\ &= (\mathbf{V}^{-1/2} \mathbf{L})(\mathbf{V}^{-1/2} \mathbf{L})' + \mathbf{V}^{-1/2} \boldsymbol{\Psi} \mathbf{V}^{-1/2}\end{aligned}$$

donde  $\mathbf{V}$  representa la matriz diagonal con las varianzas de  $X_i$

- De esta forma  $\boldsymbol{\rho}$  tiene una factorización análoga a  $\text{Cov}(\mathbf{X}) = \mathbf{L} \mathbf{L}' + \boldsymbol{\Psi}$
- Es decir  $\boldsymbol{\rho} = \mathbf{L}_Z \mathbf{L}_Z' + \boldsymbol{\Psi}_Z$ , donde

$$\mathbf{L}_Z = \mathbf{V}^{-1/2} \mathbf{L} \quad \text{y} \quad \boldsymbol{\Psi}_Z = \mathbf{V}^{-1/2} \boldsymbol{\Psi} \mathbf{V}^{-1/2}$$

# Estimación por el método de máxima verosimilitud

- Entonces por la propiedad de invarianza de los EMV, el estimador de MV de  $\rho$  es

$$\begin{aligned}\hat{\rho} &= (\hat{\mathbf{V}}^{-1/2} \hat{\mathbf{L}})(\hat{\mathbf{V}}^{-1/2} \hat{\mathbf{L}})' + \hat{\mathbf{V}}^{-1/2} \hat{\Psi} \hat{\mathbf{V}}^{-1/2} \\ &= \hat{\mathbf{L}}_Z \hat{\mathbf{L}}_Z' + \hat{\Psi}_Z.\end{aligned}$$

donde  $\hat{\mathbf{V}}^{-1/2}$ ,  $\hat{\mathbf{L}}$  y  $\hat{\Psi}$  son los estimadores de máxima verosimilitud de  $\mathbf{V}^{-1/2}$ ,  $\mathbf{L}$  y  $\Psi$ , respectivamente.

- Por lo anterior, si el análisis de máxima verosimilitud se realiza sobre la matriz de correlaciones, entonces los EMV de las communalidades son

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2 \quad i = 1, 2, \dots, p$$

- La importancia de los factores se evalúa como antes, es decir, la proporción de la varianza muestral total debido al  $j$ -ésimo factor es

$$\frac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \cdots + \hat{l}_{pj}^2}{p}$$

**Nota:** para evitar mas tedio en la notación, los  $\hat{l}_{ij}$  denotan los elementos de  $\hat{\mathbf{L}}_Z$

## Comentarios

- Los EMV de  $\mathbf{L}$ ,  $\boldsymbol{\Psi}$  y  $\mu$  se obtienen optimizando numéricamente  $\mathbf{L}(\mu, \boldsymbol{\Sigma})$ , utilizando un algoritmo iterativo tipo Newton-Raphson.
- Puede ocurrir que el algoritmo converja a un máximo local de  $\mathbf{L}(\mu, \boldsymbol{\Sigma})$ , donde algunos de los términos de  $\boldsymbol{\Psi}$  sean negativos. Esta solución impropia se denomina frecuentemente *solución de Heywood*.
- Los programas existentes cambian entonces esos valores por números positivos e intentan encontrar otro máximo local, aunque el algoritmo no siempre converge.
- La estimación mediante máxima verosimilitud es invariante ante transformaciones lineales de las variables. Por tanto, el resultado de la estimación no depende – como ocurre en componentes principales- del uso de la matriz de covarianzas o de correlaciones.

## Comentarios

- Generalmente las observaciones se estandarizan y la matriz  $\mathbf{R}$  es factorizada.
- La  $\mathbf{S}$  es sustituida por  $\mathbf{R}$  en la función de verosimilitud  $L(\mu, \Sigma)$  y los estimadores de máxima verosimilitud  $\hat{\mathbf{L}}_Z$  y  $\hat{\Psi}_Z$  se obtienen mediante optimización numérica.
- Aunque la función de verosimilitud,  $L(\mu, \Sigma)$ , es apropiada para  $\mathbf{S}$  y no para  $\mathbf{R}$ , la obtención de  $\hat{\mathbf{L}}_Z$  y  $\hat{\Psi}_Z$  equivale a obtenerlos a partir de los EMV  $\hat{\mathbf{L}}$  y  $\hat{\Psi}$  a partir de  $\mathbf{S}$ , ajustando

$$\hat{\mathbf{L}}_Z = \hat{\mathbf{V}}^{-1/2} \hat{\mathbf{L}} \quad \text{y} \quad \hat{\Psi}_Z = \hat{\mathbf{V}}^{-1/2} \hat{\Psi} \hat{\mathbf{V}}^{-1/2}$$

donde  $\hat{\mathbf{V}}^{-1/2}$  es la estimación de la matriz diagonal con los recíprocos de las desviaciones estándar muestrales.



## Comentarios

- En el sentido opuesto, dadas las estimaciones  $\hat{\mathbf{L}}_{\mathbf{Z}}$  y  $\hat{\mathbf{\Psi}}_{\mathbf{Z}}$  obtenidas de  $\mathbf{R}$ , entonces los EMV resultantes de un análisis de factores sobre la matriz  $\mathbf{S}$  son

$$\hat{\mathbf{L}} = \hat{\mathbf{V}}^{1/2} \hat{\mathbf{L}}_{\mathbf{Z}} \quad \text{y} \quad \hat{\mathbf{\Psi}} = \hat{\mathbf{V}}^{1/2} \hat{\mathbf{\Psi}}_{\mathbf{Z}} \hat{\mathbf{V}}^{1/2}$$

- Un procedimiento alternativo para maximizar  $\mathbf{L}(\mu, \Sigma)$  es considerar los factores como valores ausentes y aplicar el *algoritmo EM*.
- Además del método de máxima verosimilitud se han propuesto otros métodos aproximados para estimar  $\mathbf{L}$  y  $\mathbf{\Psi}$ , por ejemplo el método de mínimos cuadrados generalizados, mínimos cuadrados no ponderados, etc.

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

- El supuesto de normalidad permite realizar pruebas sobre la suficiencia del modelo. Es decir, se puede probar si el número de factores  $m$  es suficiente para explicar la covarianza observada.
- Suponiendo que el modelo de  $m$  factores comunes, tiene buen ajuste. Entonces  $\Sigma = LL' + \Psi$ , y probar el ajuste del modelo de  $m$  factores comunes es equivalente a probar:

$$H_0 : \underset{(p \times p)}{\Sigma} = \underset{(p \times m)}{L} \underset{(m \times p)}{L'} + \underset{(p \times p)}{\Psi}$$
$$H_1 : \underset{(p \times p)}{\Sigma} \neq \underset{(p \times m)}{L} \underset{(m \times p)}{L'} + \underset{(p \times p)}{\Psi}$$

$H_1 : \Sigma$  es cualquier otra matriz definida positiva

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

- Cuando  $\Sigma$  no tiene una forma especial, el máximo de la función de verosimilitud es proporcional a

$$|\mathbf{S}_n|^{-n/2} e^{-np/2}$$

- Bajo  $H_0$ ,  $\Sigma$  está restringida a tomar la forma  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ . En este caso el máximo de la función de verosimilitud es proporcional a

$$\begin{aligned} & \left| \hat{\Sigma} \right|^{-n/2} \exp \left( -\frac{1}{2} \text{tr} \left[ \hat{\Sigma}^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right] \right) \\ &= \left| \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi} \right|^{-n/2} \exp \left( -\frac{1}{2} n \text{tr} \left[ \left( \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi} \right)^{-1} \mathbf{S}_n \right] \right) \end{aligned}$$

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

## Resultado

- Cuando  $n$  es grande, bajo  $H_0$ , el cociente de verosimilitud

$$-2\ln\Lambda = -2\ln\left[\frac{\text{verosimilitud maximizada bajo } H_0}{\text{verosimilitud maximizada}}\right]$$

sigue aproximadamente una  $\chi^2_{v-v_0}$ , donde

- $v$  = la dimensión del espacio paramétrico de  $H_1$
- $v_0$  = la dimensión del espacio paramétrico de  $H_0$

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

- El estadístico de la razón de verosimilitud para probar  $H_0$ , es

$$\begin{aligned} -2\ln\Lambda &= -2\ln\left[\frac{\text{verosimilitud maximizada bajo } H_0}{\text{verosimilitud maximizada}}\right] \\ &= -2\ln\left(\frac{|\hat{\Sigma}|}{|\mathbf{S}_n|}\right)^{-n/2} + n\left[\text{tr}\left(\hat{\Sigma}^{-1}\mathbf{S}_n\right) - p\right] \end{aligned}$$

Donde

$$\begin{aligned} v &= p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2} \\ v_0 &= pm + p - m(m-1)/2 \end{aligned}$$

- Entonces los grados de libertad son

$$v - v_0 = \frac{1}{2}p(p+1) - \left[p(m+1) - \frac{1}{2}m(m-1)\right] = \frac{1}{2}[(p-m)^2 - (p+m)]$$

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

- Se puede probar que  $tr(\hat{\Sigma}^{-1} \mathbf{S}_n) - p = 0$ , si  $\hat{\Sigma} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}$  es el EMV de  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ . Entonces el cociente de verosimilitud se reduce a

$$-2 \ln \left( \frac{|\hat{\Sigma}|}{|\mathbf{S}_n|} \right)^{-n/2} = n \ln \left( \frac{|\hat{\Sigma}|}{|\mathbf{S}_n|} \right)$$

- Bartlett mostró que la aproximación de la ji-cuadrada a la distribución muestral de  $-2 \ln \Lambda$  mejora si se reemplaza  $n$  en la expresión anterior por el término

$$n - 1 - \frac{2p + 4m + 5}{6}$$

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

- Usando la corrección de Bartlett, se rechaza  $H_0$  al nivel de significancia  $\alpha$  si

$$(n-1-(2p+4m+5)/6) \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{[(p-m)^2 - p - m]/2}(\alpha)$$

siempre que  $n$  y  $n-p$  sean grandes.

- Como el número de grados de libertad  $1/2[(p-m)^2 - (p+m)]$  debe ser positivo, entonces  $m$  debe satisfacer

$$m < 1/2 \left( 2p+1 - (8p+1)^{1/2} \right)$$

para poder aplicar la prueba.

# Prueba de hipótesis para el número de factores comunes (con muestra grande)

- Generalmente esta prueba se aplica secuencialmente: se estima el modelo con un valor pequeño de  $m$ ,  $m = m_1$  y se prueba  $H_0$ . Si se rechaza se reestima con  $m = m_1 + 1$ , continuando hasta aceptar  $H_0$ .
- Con esta prueba se verifica la suficiencia del modelo de  $m$  factores comunes, comparando las varianzas generalizadas  $|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|$  y  $|\mathbf{S}_n|$ . Si  $n$  es grande y  $m$  es pequeño con respecto a  $p$ , la hipótesis  $H_0$  usualmente se rechazará sugiriendo agregar mas factores comunes.
- Sin embargo,  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$  podría estar suficientemente cercano a  $\mathbf{S}_n$ , de tal forma que agregar mas factores no mejoran los resultados, a pesar de que esos factores son significativos.
- Por tanto, se deben hacer otras valoraciones en la elección de  $m$ , además de los resultados de esta prueba.



# Factor Scores

- En análisis de factores, el interés se centra en la estimación de los parámetros del modelo  $\mathbf{L}$  y  $\Psi$ .
- En muchos análisis de factores puede requerirse la estimación de los  $n$  *vectores de puntajes factoriales* (*factor scores*).
- Estas cantidades frecuentemente son usadas para propósito de diagnóstico o como entrada para un análisis posterior.
- *Los factor scores* no son estimaciones de los parámetros desconocidos en el sentido usual.
- Son estimaciones de los valores para los vectores factoriales aleatorios no observados  $\mathbf{F}_j$ ,  $j = 1, \dots, n$ .
- Esto es, los factor scores son

$\hat{\mathbf{f}}_j$  = estimación de los valores  $\mathbf{f}_j$  obtenidos de  $\mathbf{F}_j$  (en el  $j$ -ésimo caso)

# Factor Scores

- La estimación es complicada por el hecho de que las cantidades no observadas  $f_j$  y  $\varepsilon_j$  son mayor en número que las observaciones  $X_j$ .
- Se han propuesto varias metodologías para estimar los factor scores, que resuelven esta problemática.
- Los métodos mas consistentes y utilizados son el de mínimos cuadrados ponderados (Bartlett) y el método por regresión (Thompson)
- El método por mínimos cuadrados ponderados supone que el vector de valores de los factores para cada observación es un parámetro a estimar
- El método por regresión supone que estos vectores son variables aleatorias

## Elementos comunes de ambos procedimientos de estimación:

- 1 Tratan los valores estimados para las cargas factoriales  $\hat{l}_{ij}$  y para las varianzas específicas  $\hat{\psi}_i$  como si fueran los valores reales.
- 2 Consideran transformaciones lineales de los datos originales (los datos se centran o se estandarizan). Generalmente se utilizan las estimaciones de las cargas rotadas, en lugar de estimaciones de las cargas originales, para calcular los factor scores. Las formulas obtenidas de los factor scores no cambian cuando las cargas rotadas son sustituidas por las cargas originales, es decir no habrá diferencia entre ellas

# Factor Scores. Estimación por Mínimos Cuadrados Ponderados

- Se asume que el vector de medias  $\mu$ , la matriz de cargas  $L$  y la matriz de varianzas específicas  $\Psi$  son conocidas para el modelo factorial

$$X - \mu = LF + \varepsilon$$

- El modelo se puede considerar un modelo de regresión donde los factores específicos  $\varepsilon$  se tratan como errores con varianzas conocidas  $\Psi$  no necesariamente iguales, donde el parámetro a estimar es el vector de factores comunes  $F$ .
- Debido a que las varianzas de los errores  $\Psi$  no son iguales, Bartlett sugirió usar mínimos cuadrados ponderados para estimar los valores de los factores comunes  $F$

# Factor Scores-Estimación por Mínimos Cuadrados Ponderados

- La suma de los cuadrados de los errores, ponderados por el recíproco de sus varianzas es

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_i} = \varepsilon' \Psi^{-1} \varepsilon = (\mathbf{x} - \mu - \mathbf{L}\mathbf{f})' \Psi^{-1} (\mathbf{x} - \mu - \mathbf{L}\mathbf{f})$$

- Bartlett propuso elegir el estimador  $\hat{\mathbf{f}}$  de  $\mathbf{f}$  que minimice la expresión anterior. La solución para  $\hat{\mathbf{f}}$  esta dada por

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

# Factor Scores. Estimación por Mínimos Cuadrados Ponderados

- Tomando los estimadores  $\hat{\mathbf{L}}, \hat{\Psi}$  y  $\hat{\mu} = \bar{\mathbf{x}}$  como los valores verdaderos de  $\mathbf{L}, \Psi$  y  $\mu$ , respectivamente, se obtienen los factor scores para el  $j$ -ésimo caso como

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

- Cuando  $\hat{\mathbf{L}}$  y  $\hat{\Psi}$  son obtenidos mediante máxima verosimilitud, estos deben satisfacer la condición de unicidad

$$\hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}} = \hat{\Delta}$$

donde  $\hat{\Delta}$  es una matriz diagonal.

# Factor Scores. Estimación por mínimos cuadrados ponderados a partir de estimadores de MV

- Los factor scores  $\hat{\mathbf{f}}_j$  obtenidos por mínimos cuadrados ponderados a partir de los estimadores de MV  $\hat{\mathbf{L}}$  y  $\hat{\mathbf{\Psi}}$  son

$$\begin{aligned}\hat{\mathbf{f}}_j &= (\hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} (\mathbf{x}_j - \hat{\boldsymbol{\mu}}) \\ &= \hat{\boldsymbol{\Delta}}^{-1} \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad j = 1, 2, \dots, n\end{aligned}$$

- Si la matriz de correlaciones es factorizada:

$$\begin{aligned}\hat{\mathbf{f}}_j &= (\hat{\mathbf{L}}_z' \hat{\mathbf{\Psi}}_z^{-1} \hat{\mathbf{L}}_z)^{-1} \hat{\mathbf{L}}_z' \hat{\mathbf{\Psi}}_z^{-1} \mathbf{z}_j \\ &= \hat{\boldsymbol{\Delta}}_z^{-1} \hat{\mathbf{L}}_z' \hat{\mathbf{\Psi}}_z^{-1} \mathbf{z}_j \quad j = 1, 2, \dots, n\end{aligned}$$

donde  $\mathbf{z}_j = \mathbf{D}^{-1/2} (\mathbf{x}_j - \bar{\mathbf{x}})$  y  $\hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}_z \hat{\mathbf{L}}_z' + \hat{\mathbf{\Psi}}_z$

# Factor scores. Estimación por mínimos cuadrados ponderados a partir de estimadores de MV

- Los factor scores obtenidos mediante el método de mínimos cuadrados ponderados a partir de los estimadores de máxima verosimilitud, tienen vector de medias muestrales igual a cero y covarianza muestral igual a cero.
- Si las cargas rotadas  $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$  se usan en lugar de las cargas originales  $\hat{\mathbf{L}}$ , los factor scores obtenidos,  $\hat{\mathbf{f}}_j^*$ , están relacionados a  $\hat{\mathbf{f}}_j$  mediante

$$\hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j, j = 1, \dots, n$$



# Factor Scores. Estimación por mínimos cuadrados (cargas estimadas mediante PCA)

- Si las cargas factoriales son estimadas por el método de componentes principales, usualmente los factor scores se obtienen usando un procedimiento de mínimos cuadrados sin ponderar.
- Implícitamente esto equivale a asumir que las  $\psi_i$  son iguales o aproximadamente iguales entre si
- Los factor scores tienen la forma

$$\hat{f}_j = (\hat{\mathbf{L}}' \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' (\mathbf{x}_j - \bar{\mathbf{x}})$$

o para datos estandarizados

$$\hat{f}_j = (\hat{\mathbf{L}}_z' \hat{\mathbf{L}}_z)^{-1} \hat{\mathbf{L}}_z' \mathbf{z}_j$$

Ahora bien, debido a que

$$\tilde{\mathbf{L}} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 & \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \end{bmatrix}$$

# Factor Scores. Estimación por mínimos cuadrados (cargas estimadas mediante PCA)

- Entonces:

$$\hat{\mathbf{f}}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\mathbf{e}}_1' (\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{\mathbf{e}}_2' (\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{\mathbf{e}}_m' (\mathbf{x}_j - \bar{\mathbf{x}}) \end{bmatrix}$$

- Los  $\hat{\mathbf{f}}_j$  son los primeros  $m$  componentes principales escalados evaluados en  $\mathbf{x}_j$
- Para estos factor scores se tiene que:

$$\frac{1}{n} \sum_{j=1}^n \hat{\mathbf{f}}_j = \mathbf{0} \quad (\text{media muestral})$$

$$\frac{1}{n-1} \sum_{j=1}^n \hat{\mathbf{f}}_j \hat{\mathbf{f}}_j' = \mathbf{I} \quad (\text{covarianza muestral})$$

# Obtención de los factor scores por el método de regresión

- Considerando el modelo original de factores

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}$$

- Asumiendo que  $\mathbf{L}$  y  $\boldsymbol{\Psi}$  son conocidas, podemos interpretar este modelo como un modelo de regresión.
- Cuando los factores comunes  $\mathbf{F} \sim \mathbf{N}_m(\mathbf{0}, \mathbf{I})$  y  $\boldsymbol{\varepsilon} \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Psi})$ , entonces la combinación lineal  $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}$  sigue una distribución  $\mathbf{N}_p(\mathbf{0}, \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})$ .
- Además, la distribución conjunta de  $\mathbf{X} - \boldsymbol{\mu}$  y  $\mathbf{F}$  es  $\mathbf{N}_{m+p}(\mathbf{0}, \boldsymbol{\Sigma}^*)$  donde

$$\boldsymbol{\Sigma}_{(m+p) \times (m+p)}^* = \left[ \begin{array}{c|c} \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} & \mathbf{L} \\ \hline \mathbf{L}' & \mathbf{I} \end{array} \right]$$

$\mathbf{0}$  es vector de ceros de tamaño  $m + p$ .

# Obtención de los factor scores por el método de regresión

- Por otro resultado de la distribución normal multivariada, la distribución condicional de  $\mathbf{F}|\mathbf{x}$  es normal multivariada con:

$$E(\mathbf{F}|\mathbf{x}) = \mathbf{L}'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

$$\text{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}\mathbf{L}$$

- Las cantidades  $\mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}$  son los coeficientes en una regresión (multivariada) de los factores sobre las variables.
- Por tanto, estimaciones de estos coeficientes producen los factor scores que son análogos a la estimación de los valores de la media condicional en un análisis de regresión multivariada.

$$\hat{\mathbf{f}} = \hat{\boldsymbol{\beta}}(\mathbf{x} - \hat{\boldsymbol{\mu}}) \quad \boldsymbol{\beta} = \mathbf{L}'(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi})^{-1}$$

# Obtención de los factor scores por el método de regresión

- Por tanto, dado cualquier vector de observaciones  $\mathbf{x}_j$ , y tomando los EMV  $\hat{\mathbf{L}}$  y  $\hat{\mathbf{\Psi}}$  y  $\hat{\mu} = \bar{\mathbf{x}}$  como los verdaderos valores, el  $j$ -ésimo factor scores está dado por:

$$\hat{f}_j = \hat{\mathbf{L}}' \hat{\mathbf{\Sigma}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) = \hat{\mathbf{L}}' (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad j = 1, 2, \dots, n$$

- El cálculo del  $\hat{f}_j$  anterior se puede simplificar usando la siguiente identidad

$$\hat{\mathbf{L}}' (\hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})^{-1} = (\mathbf{I} + \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \hat{\mathbf{\Psi}}^{-1}$$

- Esta identidad permite comparar los factor scores generados mediante regresión y mediante mínimos cuadrados ponderados.

# Obtención de los factor scores por el método de regresión

Denotamos al primero como  $\hat{f}_j^R$  y al segundo como  $\hat{f}_j^{LS}$ . Entonces de la identidad anterior:

$$\begin{aligned}\hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1} &= (I + \hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \hat{L}'\hat{\Psi}^{-1} \\ \Rightarrow \hat{L}'\hat{\Psi}^{-1} &= (I + \hat{L}'\hat{\Psi}^{-1}\hat{L}) \hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1}\end{aligned}$$

Por tanto

$$\begin{aligned}\hat{f}_j^{LS} &= (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \hat{L}'\hat{\Psi}^{-1}(x_j - \bar{x}) \\ &= (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} (I + \hat{L}'\hat{\Psi}^{-1}\hat{L}) \hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1} (x_j - \bar{x}) \\ &= \left( I + (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \right) \hat{L}' (\hat{L}\hat{L}' + \hat{\Psi})^{-1} (x_j - \bar{x}) \\ &= \left( I + (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} \right) \hat{f}_j^R\end{aligned}$$

Para los EMV,  $(\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1} = \hat{\Delta}^{-1}$  y si los elementos de  $\hat{\Delta}^{-1}$  son cercanos a cero, los métodos de regresión y mínimos cuadrados ponderados darán aproximadamente los mismos factor scores.

# Obtención de los factor scores por el método de regresión

- Una práctica común es utilizar la matriz de varianzas y covarianzas muestral  $\mathbf{S}$  en lugar de la matriz de covarianzas estimada por máxima verosimilitud  $\hat{\Sigma} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}$ . Entonces los factor scores se pueden expresar como:

$$\hat{f}_j = \hat{\mathbf{L}}' \mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \quad j = 1, 2, \dots, n$$

- O si la matriz de correlaciones es factorizada

$$\hat{f}_j = \hat{\mathbf{L}}_Z' \mathbf{R}^{-1} \mathbf{z}_j \quad j = 1, 2, \dots, n$$

donde

$$\mathbf{z}_j = \mathbf{V}^{-\frac{1}{2}}(\mathbf{x}_j - \bar{\mathbf{x}}) \quad \hat{\rho} = \hat{\mathbf{L}}_Z \hat{\mathbf{L}}_Z' + \hat{\Psi}_Z$$

- Si las cargas rotadas  $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$  se usan en lugar de las cargas originales, los factor scores  $\hat{f}_j^*$  están relacionados a  $\hat{f}_j$  por

$$\hat{f}_j^* = \mathbf{T}' \hat{f}_j \quad j = 1, 2, \dots, n$$

# Perspectivas y estrategias para el análisis de factores

- Tal vez la decisión más importante en el análisis de factores es la elección de  $m$ .
- Se puede utilizar la prueba del cociente de verosimilitud siempre y cuando la muestra sea grande y los datos se distribuyan aproximadamente normal.
- Además, la prueba rechazará con toda seguridad el modelo para  $m$  pequeño, si el número de variables y observaciones es grande.
- Sin embargo, lo ideal es obtener precisamente un modelo con un número pequeño de factores, que es cuando el AF nos proporciona una solución útil.



# Perspectivas y estrategias para el análisis de factores

- En la mayoría de los casos, la elección de  $m$  se basa en una combinación de:
  - 1 La proporción de la varianza muestral explicada por los factores,
  - 2 el conocimiento que se tenga de la situación real
  - 3 qué tan razonables son los resultados
- La elección del método de solución y el tipo de rotación son decisiones menos cruciales.
- Hasta la fecha, análisis de factores mantiene un sabor de arte, y no existe una estrategia sencilla que se pueda seguir.

# Perspectivas y estrategias para el análisis de factores

Una estrategia sugerida es la siguiente:

- ❶ Realizar un análisis de factores por el método de *componentes principales*. Este método es apropiado para una primera revisión de los datos.
  - Buscar observaciones sospechosas, graficando los factor scores.
  - Probar una rotación varimax.
- ❷ Realizar un análisis de factores por *máxima verosimilitud*, incluyendo una rotación varimax.
- ❸ Comparar las soluciones obtenidas de ambos análisis.
- ❹ Repetir los primeros tres pasos para otros números de factores comunes  $m$ . Los factores adicionales necesariamente contribuyen a la comprensión e interpretación de los datos?
- ❺ Para conjuntos grandes de datos, dividirlos aleatoriamente a la mitad y realizar un análisis de factores con cada una de las partes. Compara los dos resultados y con los obtenidos de los datos completos para revisar la estabilidad de la solución