

TÉCNICAS ESTADÍSTICAS MULTIVARIADAS PARA LA EXPLORACIÓN Y REDUCCIÓN DE DATOS

Centro de Investigación en Matemáticas A.C.

www.cimat.mx

AUTORES

M. en E. Sergio Martín Nava Muñoz
nava@cimat.mx

Dr. Jorge Raúl Pérez Gallardo
raul.perez@cimat.mx

Centro de Investigación en Matemáticas, A.C.
Unidad Aguascalientes
MÉXICO
2018



Unidad III Análisis de Factores Comunes

Antes de empezar:

A lo largo del material de estudio usted encontrará referencias visuales que le indicarán el tipo de texto que está leyendo.

A continuación, se las presentamos.



Importante:

Presentan definiciones en relación a los conceptos trabajados.



Ejemplos:

Con este ícono se destacan ejemplos o casos.



Integrando Ideas:

Son párrafos que sintetizan las ideas desarrolladas hasta ese momento.



Actividades optativas y sin entrega obligatoria:

Indican todas las actividades que usted podrá realizar a lo largo de cada unidad y que cuentan con la respuesta al final de la misma.



Es el ícono del software estadístico que se utilizará en este curso. Su presencia indica que es el momento de utilizarlo.

Objetivos:

Durante el transcurso de esta unidad se pretende que el participante logre:

- Diferenciar la técnica de Análisis de Factores de otras técnicas multivariantes.
- Identificar el tipo de casos en que es conveniente el uso de la técnica de Análisis Factorial.
- Identificar las diferencias principales entre la técnica de Análisis de Componentes Principales y los modelos de Análisis de Factores.
- Comprender el concepto de rotación de factores.
- Determinar las restricciones que tiene el Análisis de Factores.

CONTENIDO

1.	INTRODUCCIÓN	4
2.	RELACIÓN CON EL ANÁLISIS DE COMPONENTES PRINCIPALES	5
3.	MODELO UNIFACTORIAL O DE ÚNICO FACTOR	6
4.	MODELO PARA K-FACTORES O MULTIFACTORIAL	8
5.	ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO DE K-FACTORES.....	9
5.1.	NÚMERO MÁXIMO DE FACTORES COMUNES.....	10
5.2.	MÉTODO DEL FACTOR PRINCIPAL.....	11
5.2.1.	<i>Estimación de las communalidades.....</i>	12
5.3.	MÉTODO DE LA MÁXIMA VEROSIMILITUD.....	13
5.3.1.	<i>Hipótesis sobre el número de factores.....</i>	14
6.	ROTACIÓN DE LOS FACTORES	15
6.1.	ROTACIÓN ORTOGONAL	16
6.2.	ROTACIÓN OBLICUA.....	17
7.	CRITERIOS PARA LA SIGNIFICANCIA DE LAS CARGAS FACTORIALES...	17
8.	ESTIMACIÓN DE LAS PUNTUACIONES DE LOS FACTORES.....	18
9.	VALIDACIÓN DEL MODELO RESULTANTE.....	19

Frecuentemente en las ciencias encargadas del estudio de comportamiento humano como la psicología no es posible medir de forma directa el concepto de interés del estudio, por ejemplo la *inteligencia* o la *clase social*. En estos dos casos, el investigador o el analista están obligado a examinar estos conceptos de forma indirecta mediante la información que puede proporcionar un conjunto de variables que pueden ser medidas y observadas directamente y que pueden ser consideradas como indicadores de los conceptos que se está interesado estudiar. El psicólogo se ve atraído por evaluar la *inteligencia* de las personas y para ello puede estudiar las calificaciones registradas de diferentes pruebas cognoscitivas en la espera que esas calificaciones dependen de alguna forma de lo que se conoce como inteligencia. Para conocer la *clase social* de las personas, un sociólogo puede hacer preguntas sobre la ocupación de las personas, su trayectoria académica, las propiedades que posee, etc., en el supuesto que las variables medidas reflejen el concepto *clase social*. El economista puede estar interesado en medir el *desarrollo* de un país mediante un pequeño conjunto de variables económicas, sociales y demográficas que asume se relacionan con el desarrollo.

Estas variables, *inteligencia*, *clase social* o *desarrollo*, son conceptos que no pueden ser medidos de forma directa pero puede asumirse que están relacionados con cierto número de variables medibles. El método más empleado para ayudar a descubrir la relación entre ambos tipos de variables es el *Análisis Factorial* o de *Factores Comunes* (AF).

1. Introducción



El **Análisis de Factores** (AF) es un método multivariante que pretende expresar p variables observables o manifiestas como una combinación lineal de m variables hipotéticas o latentes, a las que se denominan *factores*.

El modelo que se basa este método es parecido al de una *Regresión Múltiple* solo que las variables independientes se relaciona con una o varias variables *latentes* o no medibles. De la misma forma, la estimación de los coeficientes de regresión (denominados en AF como *cargas factoriales*) no es posible realizarla de forma directa.

El Análisis de Factores se divide en dos categorías. El *Análisis Factorial Exploratorio* es empleado para investigar la relación entre las variables manifiestas y los factores sin realizar supuestos *a priori* sobre que variables manifiestas están relacionada con determinado factor. El *Análisis Factorial Confirmatorio* se utiliza para evaluar si un determinado modelo factorial definido *a priori* provee una explicación satisfactoria de la matriz de covarianza o correlación entre las variables manifiestas. Esta unidad estará dedicada a explicar la técnica de *Análisis Factorial Exploratorio*.

El Análisis de Factores considerando un único factor fue introducido por el interés de Karl Pearson y de Charles Spearman en comprender las dimensiones de la *inteligencia humana* en la década de los 30s, por lo que muchos de los avances de esta técnica tienen una fuerte relación con el área de la psicometría. Aunque en la segunda mitad del siglo XX se le ha dado un enfoque más amplio al desarrollarse diversas metodologías estadísticas. Esto ha hecho extensa su aplicación en otras disciplinas como relaciones internacional, comunicaciones, taxonomía, biología, medicina, meteorología y educación.

2. Relación con el Análisis de Componentes Principales

El Análisis de Factores sigue una formulación muy parecida al *Análisis de Componentes Principales* (ACP) ya que ambos buscan explicar un conjunto de datos multivariados en un número reducido de dimensiones pero existen ciertas diferencias en procedimiento seguido. La primera diferencia que a destacar es que los componentes principales se construyen para dar una explicación a la varianza total que presenta el conjunto de variables bajo estudio, mientras que los factores son construidos para buscar alguna explicación a las covarianzas o correlaciones existente entre las variables mediante un reducido conjunto de factores en común.

Una segunda diferencia es que si se agrega un componente principal adicional a los retenidos, es decir pasar de m componentes a $m + 1$ componentes, las cargas de los m componentes no cambian. En el caso del AF las cargas de todos los factores se ven alteradas si el número de factores retenidos aumenta.

Otra diferencia que se presenta es que el ACP es una herramienta sencilla más orientada a un análisis descriptivo mientras que el AF presupone un modelo estadístico formal en donde la estimación de las cargas factoriales es más compleja.

A pesar de las diferencias mencionadas anteriormente los resultados arrojados por ambas técnicas son casi siempre similares cuando la variabilidad es pequeña. Sin embargo, cuando la varianza total es grande, esta será absorbida por todos los componentes principales, mientras que el AF hace una distinción especial.

Un punto central que ambas técnicas comparten es que las variables que se observan deben estar *correlacionadas*. Si las variables no están correlacionadas se ve reflejado en que en el AF no hay nada que explicar y en el ACP los componentes son similares a las variables originales.

3. Modelo unifactorial o de único factor

Partiendo del trabajo que Spearman realizó sobre la *inteligencia* de los niños mediante las calificaciones obtenidas en tres materias, Clásicos (X_1), Francés (X_2) e Inglés (X_3) se ilustra este primer modelo factorial. La matriz \mathbf{R} contiene los coeficientes de correlación de estas tres variables:

$$\mathbf{R} = \begin{pmatrix} 1.00 & & \\ 0.83 & 1.00 & \\ 0.78 & 0.67 & 1.00 \end{pmatrix}$$

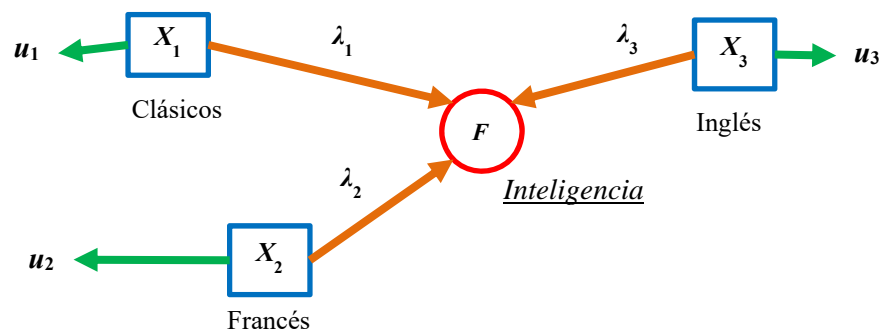
De la matriz anterior se observa una fuerte correlación entre las tres variables por lo que el uso de un modelo de AF es pertinente. Si se considera que existe un único *factor* o *variable latente* que relaciona estas tres variables observadas, el modelo factorial resultante sería el siguiente:

$$\text{Clásicos} = X_1 = \lambda_1 f + u_1,$$

$$\text{Francés} = X_2 = \lambda_2 f + u_2,$$

$$\text{Inglés} = X_3 = \lambda_3 f + u_3$$

El modelo representa una regresión simple de cada una de las variables observadas con un único factor en común, f . El factor f en este ejemplo representa la variable *inteligencia* o *habilidad intelectual* que se buscaba medir en un inicio. Los términos λ_1 , λ_2 y λ_3 que representan los coeficientes de regresión, que en el contexto de AF son conocidos como *cargas factoriales*, mientras que los términos u_1 , u_2 y u_3 representan el término de perturbación aleatoria y tendrán una varianza pequeña si la variable observada a la que está asociada está fuertemente relacionada con la *variable latente* o *factor*. La figura 1 muestra las relaciones anteriores.



Matemáticamente, el modelo factorial con un único factor quedaría definido de la siguiente forma:



Asumiendo un conjunto de variables observadas o manifiestas $\mathbf{X}^T = (X_1, X_2, \dots, X_p)$, presumiblemente relacionadas con una única variable latente f , el modelo factorial es:

$$\mathbf{X} = \mathbf{\Lambda}f + \mathbf{u}$$

donde:

1. \mathbf{X} es un vector $(p \times 1)$ de variables observadas de una población de interés.
2. $\mathbf{\Lambda}$ es un vector $(p \times 1)$ de valores constantes desconocidos. Estos valores representan las cargas factoriales que describe como el factor f afecta a las variables observadas \mathbf{X} .
3. \mathbf{u} es un vector $(p \times 1)$ de perturbaciones no observadas. Recoge el efecto de todo aquello distinto al factor que afecta a \mathbf{X} . Se asume que \mathbf{u} tiene una distribución $N(0, \boldsymbol{\psi})$ donde $\boldsymbol{\psi}$ es diagonal. Estas perturbaciones están incorrelacionadas con el factor y se denomina *variación específica*.

De lo anterior se desprende que:

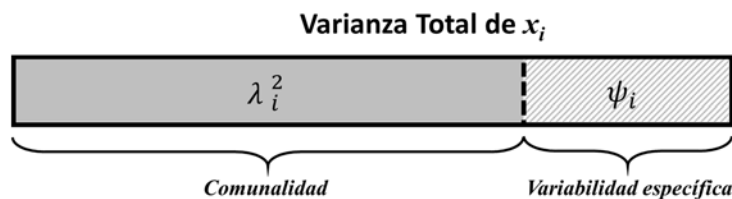
- I. El modelo asume que las variables observadas tienen una media igual a cero.
- II. El factor f contiene la parte de la variabilidad que es común a todas las variables.
- III. Cada X_i está influida por un factor único u_i que representa la parte de la variabilidad que no se puede explicar por medio del factor.
- IV. Las cargas factoriales representan la correlación existente entre las *variables observadas* y el *factor*, $\text{corr}(X_i, f) = \lambda_i$

Debido a que no se puede observar las unidades de medición del factor, éstas son arbitrarias y puede asumirse que tienen una forma estandarizada con media cero y desviación estándar de 1.

El modelo de AF presupone que la varianza de la variable X_i está dado por:

$$\sigma_i^2 = \lambda_i^2 + \psi_i \quad i = 1, \dots, p$$

Donde ψ_i es la varianza de u_i . Por lo tanto, el AF reconoce que la varianza de cada una de las variables puede ser separada en dos partes: la primera llamada *comunalidad*, $h_i^2 = \lambda_i^2$, que representa la varianza que comparte con las otras variables vía el factor común. La segunda parte, ψ_i , llamada *varianza única* o *específica* y está relacionada con la variabilidad de X_i que no comparte con las otras vía el factor. La figura 2 resume lo anterior.



Adicionalmente, la covarianza entre las variables X_i y X_j resulta ser:

$$\sigma_{ij} = \lambda_i \lambda_j \quad i \neq j = 1, \dots, p$$

Como se aprecia en el término anterior los factores solo buscan cuantificar la relación existente entre las variables observadas.

4. Modelo para k-factores o multifactorial

Supóngase ahora el caso que se tiene la necesidad de medir dos conceptos tales como *calidad de la comida* y *calidad del servicio* que ofrece un restaurante. Como en el caso de la *inteligencia* no es posible hacer una medición directa, por lo que estos dos conceptos son evaluados indirectamente mediante otras variables como: *sabor de la comida*, *tiempo de espera*, *limpieza del lugar*, *amabilidad del personal*, *presentación del platillo*, o *frescura de los alimentos*.



El modelo del **Análisis de Factores para k factores comunes**, asume que el conjunto de variables observadas o manifiestas $\mathbf{X}^T = (X_1, X_2, \dots, X_p)$ dependen de k variables latentes o factores f_1, f_2, \dots, f_k , donde $k < p$, siendo p las perturbaciones únicas, como se muestra en el modelo lineal siguiente:

$$\begin{aligned} X_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \dots + \lambda_{1k} f_k + u_1, \\ X_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \dots + \lambda_{2k} f_k + u_2, \\ &\vdots \\ X_p &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \dots + \lambda_{pk} f_k + u_p \end{aligned}$$

De manera más concisa lo anterior podría describirse como:

$$\mathbf{X} = \mathbf{\Lambda} \mathbf{f} + \mathbf{u}$$

donde:

1. \mathbf{f} es un vector $(k \times 1)$ de variables latentes o factores no observados. Se asume que siguen una distribución $N(0, I)$.
2. $\mathbf{\Lambda}$ es un vector $(p \times k)$ de valores constantes desconocidos ($p < k$). Estos valores representan las *cargas factoriales* que describen como los k -factores, \mathbf{f} , afectan a las variables observadas \mathbf{X} .
3. \mathbf{u} es un vector $(p \times 1)$ de perturbaciones no observadas. Recoge el efecto de todo aquello distinto al factor que afecta a \mathbf{X} . Se asume que \mathbf{u} tiene una distribución $N(0, \psi)$ donde ψ es diagonal.

El modelo anterior considera que tanto los factores comunes \mathbf{f} , y las perturbaciones \mathbf{u} no están correlacionados internamente ni entre ellos.

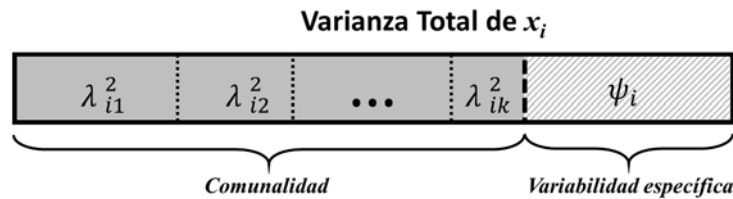
$$\begin{aligned} \text{corr}(f_i, f_j) &= 0, \quad i \neq j = 1, \dots, k \\ \text{corr}(u_i, u_j) &= 0, \quad i \neq j = 1, \dots, p \end{aligned}$$

$$\text{corr}(f_i, u_j) = 0, \quad i = 1, \dots, k \quad j = 1, \dots, p$$

El modelo unifactorial se presenta cuando el valor de $k = 1$. La varianza de la variable X_i queda definida de la siguiente forma:

$$\sigma_i^2 = \sum_{j=1}^k \lambda_{ij}^2 + \psi_i \quad i = 1, \dots, p$$

Consecuentemente, la *comunalidad* de la variable está representada por el término $\sigma_i^2 = \sum_{j=1}^k \lambda_{ij}^2$ y la *varianza única* por ψ_i . La figura 3 muestra esta relación.



Adicionalmente la expresión de covarianza es:

$$\sigma_{ij} = \sum_{l=1}^k \lambda_{il} \lambda_{jl} \quad i \neq j = 1, \dots, p$$

El resultado anterior muestra que el modelo de análisis de k -factores determina que la matriz de covarianza de la población, Σ , de las variables observadas tiene la forma

$$\Sigma = \Lambda \Lambda^T + \Psi \quad \text{donde } \Psi = \text{diag}(\psi_i)$$

De forma práctica, Σ es estimada mediante la covarianza de la matriz S de la muestra por lo que solo será necesario estimar Λ y Ψ .

5. Estimación de los parámetros del modelo de k -factores

Los parámetros que se deben de estimar son las *cargas factoriales* y las *varianzas específicas*, por lo que esencialmente hay que encontrar la matriz de cargas factoriales estimadas, $\hat{\Lambda}$, y la matriz diagonal que contenga las varianzas específicas estimadas, $\hat{\Psi}$. Si se emplea la matriz de covarianza de la muestra S , entonces

$$S \approx \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi}$$

Si se conocen los valores de la matriz de cargas factoriales estimadas, es posible estimar la varianza específica como:

$$\hat{\psi}_i = s_i^2 - \sum_{j=1}^k \hat{\lambda}_{ij}^2 \quad i = 1, \dots, p$$

Para ilustrar la forma de estimar los parámetros del modelo se retoma el ejemplo del modelo de factor simple de la sección 3.

Ejemplo 5.1. Inteligencia. Lo primero que se observa es que el número de parámetros a ser estimados en el modelo, tres cargas factoriales y tres varianzas específicas, es igual al número de elementos independientes en \mathbf{R} (las tres correlaciones y las tres varianzas estandarizadas de la diagonal).

$$\begin{pmatrix} 1.00 & & \\ 0.83 & 1.00 & \\ 0.78 & 0.67 & 1.00 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} (\lambda_1 \quad \lambda_2 \quad \lambda_3) + \begin{pmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{pmatrix}$$

Las seis ecuaciones derivadas del modelo son:

$$\begin{aligned} \hat{\psi}_1 &= 1.00 - \hat{\lambda}_1^2 \\ \hat{\psi}_2 &= 1.00 - \hat{\lambda}_2^2 \\ \hat{\psi}_3 &= 1.00 - \hat{\lambda}_3^2 \\ 0.83 &= \hat{\lambda}_1 \hat{\lambda}_2 \\ 0.78 &= \hat{\lambda}_1 \hat{\lambda}_3 \\ 0.67 &= \hat{\lambda}_2 \hat{\lambda}_3 \end{aligned}$$

La solución del sistema de ecuaciones anterior es:

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

5.1. Número máximo de factores comunes

El número de factores comunes, k , está limitado por un valor máximo k_a , el cual puede ser determinado teniendo en cuenta que existen $p(p-1)/2$ correlaciones diferentes y pk cargas factoriales. Pero si $\mathbf{\Lambda}$ es una matriz factorial con factores \mathbf{f} , también lo es $\mathbf{\Lambda}^T$, con factores $\mathbf{f}^* = \mathbf{T}^T \mathbf{F}$, donde \mathbf{T} es matriz ortogonal. Como $\mathbf{T}^T \mathbf{T} = \mathbf{I}$, se consideran $k(k-1)/2$ restricciones y el número de parámetros libres de $\mathbf{\Lambda}$ será $pk - k(k-1)/2$. Por lo que el número de correlaciones menos el número de parámetros libres es:

$$d = p(p-1)/2 - [pk - k(k-1)/2] = \frac{1}{2}[(p-k)^2 - p - k]$$

Igualando d a 0 y resolviendo la ecuación de segundo grado se obtienen que

$$k \leq k_a = \frac{1}{2} [2p + 1 - \sqrt{8p + 1}]$$

Siguiendo la ecuación anterior, el número máximo de factores para 6 variables es 3. Se tiene un modelo factorial *sobredeterminado* si $k > k_a$, puesto que hay más cargas factoriales libres que correlaciones. Si $k = k_a$ el modelo es determinado y podemos encontrar $\mathbf{\Lambda}$ de forma algebraica a partir de \mathbf{R} como en el ejemplo.

Desde la estadística, resulta interesante cuando $k < k_a$, debido a la necesidad de plantear una método de estimación. El método del *factor principal* y el método de *máxima verosimilitud* son utilizados frecuentemente.

5.2. Método del factor principal

Este método es una técnica similar al empleado en el Análisis de Componentes Principales (*eigenvalores* y *eigenvectores*). Tienen la ventaja de que evitar tener que resolver las ecuaciones de *máxima verosimilitud* como se explicara más adelante en el documento. Muchos programas dedicados al análisis estadístico de datos lo tienen programado debido a su simplicidad. El principio que sigue es el siguiente: supóngase que es posible obtener una estimación inicial de la matriz de varianzas de las perturbaciones $\hat{\Psi}$. Entonces,

$$\mathbf{S} - \hat{\Psi} = \mathbf{\Lambda}\mathbf{\Lambda}^T$$

Debido a que $\mathbf{S} - \hat{\Psi}$ es simétrica, es posible descomponerse como:

$$\mathbf{S} - \hat{\Psi} = \mathbf{H}\mathbf{G}\mathbf{H}^T = (\mathbf{H}\mathbf{G}^{1/2})(\mathbf{H}\mathbf{G}^{1/2})^T$$

Donde \mathbf{H} es cuadrada de orden p y ortogonal, \mathbf{G} también es de orden p , diagonal y contiene las raíces características de $\mathbf{S} - \hat{\Psi}$. El modelo factorial establece que \mathbf{G} debe ser diagonal del tipo:

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{1k \times k} & \mathbf{O}_{k \times (p-k)} \\ \mathbf{O}_{(p-k) \times k} & \mathbf{O}_{(p-k) \times (p-k)} \end{pmatrix}$$

Puesto que $\mathbf{S} - \hat{\Psi}$ tiene rango k . Por tanto, si \mathbf{H}_1 es la matriz $p \times k$ que contiene los vectores propios asociados a los valores propios no nulos de \mathbf{G}_1 se toma como estimador de $\mathbf{\Lambda}$ la matriz $p \times k$:

$$\hat{\mathbf{\Lambda}} = \mathbf{H}_1\mathbf{G}_1^{1/2}$$

con lo que se resuelve el problema. Obsérvese que la normalización resultante es:

$$\hat{\mathbf{\Lambda}}^T \hat{\mathbf{\Lambda}} = \mathbf{G}_1^{1/2} \mathbf{H}_1^T \mathbf{H}_1 \mathbf{G}_1^{1/2} = \mathbf{G}_1 = \text{Diagonal}$$

ya que los vectores propios de matrices simétricas son ortogonales, por lo que $\mathbf{H}_1^T \mathbf{H}_1 = \mathbf{I}_k$. Por tanto, con este método se obtienen estimadores de la matriz $\hat{\mathbf{\Lambda}}$ con columnas ortogonales entre sí. En la práctica la estimación se lleva a cabo de forma iterativa como sigue:

1. Partir de una estimación inicial de $\hat{\mathbf{\Lambda}}_i$ o de $\hat{\Psi}_i$ mediante $\hat{\Psi}_i = \text{diag}(\hat{\mathbf{\Lambda}} \hat{\mathbf{\Lambda}}^T)$.
2. Calcular la matriz cuadrada y simétrica $\mathbf{Q}_i = \mathbf{S} - \hat{\Psi}_i$.
3. Obtener la descomposición espectral de \mathbf{Q}_i de forma

$$\mathbf{Q}_i = \mathbf{H}_{1i} \mathbf{G}_{1i} \mathbf{H}_{1i}^T + \mathbf{H}_{2i} \mathbf{G}_{2i} \mathbf{H}_{2i}^T$$

donde \mathbf{G}_{1i} contiene los m mayores valores propios de \mathbf{Q}_i y \mathbf{H}_{1i} sus valores propios. Se elige k de manera que los restantes vectores propios contenidos en \mathbf{G}_{2i} sean todos pequeños y de tamaño similar. La matriz \mathbf{Q}_i puede no ser definida positiva y algunos de sus valores propios pueden ser negativos. Esto no es un problema grave si estos valores propios son muy pequeños y se puede suponerlos próximos a cero.

4. Tomar $\hat{\mathbf{\Lambda}}_{i+1} = \mathbf{H}_{1i} \mathbf{G}_{1i}^{1/2}$ y volver a (1). Iterar hasta convergencia, es decir hasta que $\|\hat{\mathbf{\Lambda}}_{n+1} - \hat{\mathbf{\Lambda}}_n\| < \varepsilon$

Los estimadores obtenidos serán consistentes pero no eficientes, como en el caso de *Máxima verosimilitud*. Tampoco son invariantes ante transformaciones lineales, es decir, no se obtiene necesariamente el mismo resultado con la matriz de covarianzas y con la de correlaciones.

Para llevar a la práctica esta idea, se debe especificar cómo obtener el estimador inicial $\hat{\Psi}$, problema que se conoce como la estimación de las *comunalidades*.

5.2.1. Estimación de las comunalidades

Estimar los términos ψ_i^2 equivale a definir valores para los términos diagonales, h_i^2 , de $\mathbf{\Lambda} \mathbf{\Lambda}^T$, ya que $h_i^2 = s_i^2 - \psi_i^2$. Existen las siguientes alternativas:

1. Tomar $\hat{\Psi}_i = 0$. Esto equivale a extraer los componentes principales de \mathbf{S} . Supone tomar $h_i^2 = s_i^2$ (en el caso de correlaciones $h_i^2 = 1$) que es claramente su valor máximo, por lo que comienza con un sesgo importante.
2. Tomar $\hat{\Psi}_j^2 = 1/s_{jj}^*$, donde s_{jj}^* es el elemento diagonal j -ésimo de la matriz de precisión \mathbf{S}^{-1} . Lo anterior equivale a tomar h_j^2 como:

$$\hat{h}_j^2 = s_j^2 - s_j^2(1 - R_j^2) = s_j^2 R_j^2$$

3. donde R_j^2 es el coeficiente de correlación múltiple entre X_j y el resto de las variables. Intuitivamente, cuanto mayor sea R_j^2 mayor será la *comunalidad* \hat{h}_j^2 . Con este método se empieza con una estimación sesgada a la baja de \hat{h}_j^2 , ya que $\hat{h}_j^2 \leq h_j^2$.

5.3. Método de la máxima verosimilitud

Desde un punto de vista estadístico, el método de la *máxima verosimilitud* es quizá es más adecuado para estimar los parámetros del Análisis de Factores. La esencia de este método es que asume que los datos a analizar tienen una distribución normal multivariada $N_p(\mu, V)$. Por lo tanto, sustituyendo μ por \bar{x} la función de verosimilitud es:

$$\log(V|X) = -\frac{n}{2} [\log V + \text{traza}(SV^{-1})]$$

Sustituyendo V por $V = \Lambda\Lambda^T + \Psi$, la función de soporte de Λ y Ψ es:

$$L(\Lambda, \Psi) = -\frac{n}{2} [\log(\Lambda\Lambda^T + \Psi) + \text{traza}(S(\Lambda\Lambda^T + \Psi)^{-1})]$$

Los estimadores de máxima verosimilitud se obtienen maximizando la ecuación anterior respecto a las matrices Λ y Ψ , obteniendo:

$$\begin{aligned} \hat{\Psi} &= \text{diag}(S - \hat{\Lambda}\hat{\Lambda}^T) \\ [\hat{\Psi}^{-1/2}(S - \mathbf{I})\hat{\Psi}^{-1/2}] (\hat{\Psi}^{-1/2}\hat{\Lambda}) &= (\hat{\Psi}^{-1/2}\hat{\Lambda}) \mathbf{D} \end{aligned}$$

donde \mathbf{D} es la matriz resultado de la normalización $\hat{\Lambda}^T \hat{\Psi}^{-1} \hat{\Lambda} = \mathbf{D}$.

Estas tres ecuaciones permiten resolver el sistema utilizando un algoritmo iterativo tipo Newton-Raphson. La solución numérica es a veces difícil porque puede no haber una solución en la cual $\hat{\Psi}$ sea definida positiva, y es necesario entonces acudir a la estimación con restricciones. Obsérvese que la última ecuación conduce a una ecuación de valores propios: dice que $\hat{\Psi}^{-1/2}\hat{\Lambda}$ contiene los vectores propios de la matriz simétrica $[\hat{\Psi}^{-1/2}(S - \mathbf{I})\hat{\Psi}^{-1/2}]$ y que \mathbf{D} contiene los valores propios.

El algoritmo iterativo para resolver estas ecuaciones es:

1. Partir de una estimación inicial. Si se cuenta con una estimación $\hat{\Lambda}_i$ ($i = 1$ la primera vez), por ejemplo por el método del factor principal, se calcula la matriz $\hat{\Psi}_i$ mediante $\hat{\Psi} = \text{diag}(S - \hat{\Lambda}\hat{\Lambda}^T)$. Alternativamente, se estima la matriz $\hat{\Psi}_i$ directamente por el método del factor principal.

2. Se calcula la matriz cuadrada simétrica $\mathbf{A}_i = \hat{\Psi}_i^{-1/2} (S - \hat{\Psi}_i) \hat{\Psi}_i^{-1/2} = \hat{\Psi}_i^{-1/2} S \hat{\Psi}_i^{-1/2} - \mathbf{I}$. Esta matriz pondera los términos de S por su importancia en términos de los componentes específicos.

3. Se obtiene la descomposición espectral de \mathbf{A}_i de forma que

$$\mathbf{A}_i = \mathbf{H}_{1i} \mathbf{G}_{1i} \mathbf{H}_{1i}^T + \mathbf{H}_{2i} \mathbf{G}_{2i} \mathbf{H}_{2i}^T$$

donde los k mayores valores propios de \mathbf{A}_i están en la matriz diagonal $(k \times k)$, \mathbf{G}_{1i} y los $p - k$ menores de la \mathbf{G}_{2i} y \mathbf{H}_{1i} , y \mathbf{H}_{2i} contienen los correspondientes vectores propios.

4. Se toma $\hat{\Lambda}_{i+1} = \hat{\Psi}_i \mathbf{H}_{1i} \mathbf{G}_{1i}^{1/2}$ y se sustituye en la función de verosimilitud, que se maximiza respecto a Ψ . Esta parte es fácil de hacer con un algoritmo de optimización no lineal. Con el resultado obtenido se vuelve a (2), iterando hasta la convergencia.

5.3.1. Hipótesis sobre el número de factores

Una ventaja del método de la máxima verosimilitud es que permite formular un test de hipótesis sobre la estructura factorial y el número k de factores comunes. La prueba queda planteada de la siguiente manera:

$$H_0: \Sigma = \Lambda \Lambda^T + \Psi$$

$$H_1: \Sigma \neq \Lambda \Lambda^T + \Psi$$

Si $\hat{\Sigma} = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi}$, los máximos de logaritmo de la razón de verosimilitud son

$$H_0: -\frac{n}{2} [\log(\hat{\Sigma}) + \text{traza}(\hat{\Sigma}^{-1} S)]$$

$$H_1: -\frac{n}{2} [\log(S) + p]$$

El estadístico de prueba es

$$C_k = n [\log(\hat{\Sigma}) - \log(S) + \text{traza}(\hat{\Sigma}^{-1} S) - p]$$

El cual sigue asintóticamente la distribución ji-cuadrada con

$$n = \frac{1}{2} [(p - k)^2 - p - k]$$

grados de libertad. Lo anterior es importante debido a que cuando el tamaño muestral es grande y k es pequeño con relación a p , si los datos no siguen una distribución normal multivariante el contraste conduce generalmente a rechazar H_0 . Este es un problema frecuente en contraste de hipótesis con muestras grandes, donde se tiende a rechazar H_0 . Este contraste es muy sensible a desviaciones de la normalidad por lo que en la práctica el estadístico se utiliza como medida de ajuste del modelo más que como un test formal.

6. Rotación de los factores

Uno de los inconvenientes que se presenta el AF es que no existe una solución única para la obtención de la matriz de cargas factoriales, Λ . Lo anterior se comprueba al introducir una matriz ortogonal de orden $k \times k$ y por lo tanto se reescribe la ecuación de regresión para las variables latentes y observadas de la siguiente manera

$$\mathbf{X} = (\Lambda \mathbf{M})(\mathbf{M}^T \mathbf{f}) + \mathbf{u}$$

El “nuevo” modelo satisface los requerimientos indicados para un modelo con k -factores con nuevos factores $\mathbf{f}^* = \mathbf{M}\mathbf{f}$ y la nueva matriz de cargas factoriales $\Lambda \mathbf{M}$. Esto implica que la matriz de covarianza de las variables observadas sea

$$\Sigma = (\Lambda \mathbf{M})(\Lambda \mathbf{M})^T + \Psi$$

Dado que $\mathbf{M}\mathbf{M}^T = \mathbf{I}$, la expresión anterior se reduce a $\Sigma = \Lambda \Lambda^T + \Psi$. Por lo que los factores \mathbf{f} con cargas Λ y los factores \mathbf{f}^* con cargas $\Lambda \mathbf{M}$ son, para cualquier matriz ortogonal \mathbf{M} , equivalentes para explicar la matriz de covarianza de las variables observadas. Lo anterior hace suponer que existen un número infinito de soluciones del modelo de AF planteado.

Este problema es resuelto al restringir el modelo original. Una posibilidad es solicitar una matriz \mathbf{G} dada por

$$\mathbf{G} = \Lambda \Psi^{-1} \Lambda$$

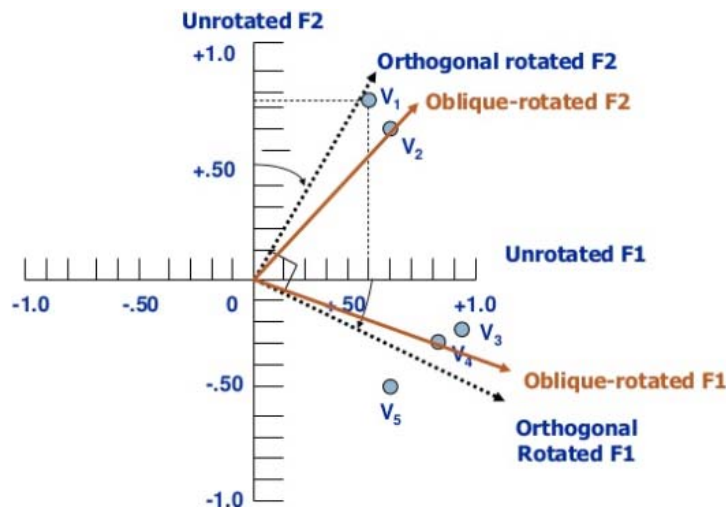
Se diagonal con sus elementos ordenados de forma decreciente en magnitud. Esta restricción obliga a que el primer factor tenga la máxima contribución de la varianza en común de las variables observadas, y el segundo factor recogerá la máxima contribución de la varianza restante cuidando que sea incorrelacionada con la primera, y así en adelante.

Esta característica resulta interesante debido a que, normalmente, la matriz obtenida no define unos factores fáciles de interpretar. Intuitivamente, será más fácil interpretar un factor cuando se asocia a un bloque de variables observadas. Esto ocurrirá si las columnas de la matriz de carga, que representan el efecto de cada factor sobre las variables observadas, contienen valores altos para ciertas variables y pequeños para otras. Esta idea puede plantearse de distintas formas que dan lugar a distintos criterios para definir la *rotación*. Los coeficientes de la matriz ortogonal \mathbf{M} que define la rotación se obtendrán minimizando una función objetivo que expresa la simplicidad deseada en la representación conseguida al rotar de tal forma que las cargas del modelo equivalente sean $\Lambda^* = \Lambda \mathbf{M}$.

La rotación de factores simplemente permitirá ofrecer una descripción más sencilla de la estructura sin alterar dicha estructura. Se han propuesto diferentes versiones sobre cómo transformar la matriz factorial a fin de obtener una estructura simple de los factores. Esencialmente se trata de conseguir que unas cargas sean altas a costa de otras, que serán bajas, para así destacar la influencia de los factores comunes sobre las variables observables. Dos tipos de rotaciones son posibles:

- *Rotación ortogonal*, en estos métodos están restringidos a que los factores rotados no presentan correlación
- *Rotación oblicua*, donde los métodos permiten cierta correlación entre factores.

Ambos métodos se aplican después de haber generado la matriz original de cargas. Cuando se emplea la *rotación ortogonal*, la matriz de correlación entre los factores después de la rotación es la matriz identidad mientras que con la *rotación oblicua*, la matriz de correlación contendrá a la unidad en la diagonal pero no tiene restricción en el valor de los elementos fuera de ella. No hay una regla establecida para decidir el tipo de rotación a emplear. Como una posible regla a tomar en cuenta es que si el analista busca principalmente que los resultados se ajusten particularmente mejor sus datos, entonces los factores deberán ser rotados de forma oblicua. Por el otro lado si busca generalizar sus trabajos y que sirvan para aplicar una técnica multi-variada por posterior, la rotación ortogonal será la más adecuada.



Esquema de rotación ortogonal y rotación oblicua

A continuación se explican algunos de los métodos más empleados.

6.1. Rotación ortogonal

- *Varimax*, propuesta por Kaiser (1958), con la intención de tener variables con cargas altas (cerca de 1 o -1) y algunas cargas cercanas a 0 en cada factor. La lógica está en que la interpretación se facilita cuando las correlaciones variable-factor están (1)

cerca de -1 o 1 , indicando una clara asociación positiva o negativa entre el factor y la variable; o cercanas a 0 señalando una clara ausencia de asociación. Este método ha demostrado tener más éxito como aproximación analítica para lograr una rotación ortogonal.

- *Quartimax*, sugerida por Carroll (1953), fuerza a que una determinada variable se correlacione altamente con un factor y poco o nada con los otros factores.

6.2. Rotación obliqua

- *Oblimin*, ideada por Jennrich y Sampson (1966), intenta encontrar una estructura simple observando el patrón que guarda la matriz de factores mediante un parámetro que controle el grado de correlación entre factores. Los valores sugeridos son -0.5 y 0.5 .
- *Promax*, propuesto por Hendrickson y White (1964), opera al incrementar las cargas de una solución ortogonal (generalmente *varimax*) a alguna potencia. El objetivo es obtener una solución que genere la mejor estructura usando el exponente más bajo posible y la correlación más baja entre factores.

7. Criterios para la significancia de las cargas factoriales

Al interpretar los factores, es necesario tomar la decisión en torno a que cargas factoriales merece la pena considerar como importantes. Para ello hay que tomar en cuenta además del número de variables, la significancia práctica y estadística que afectan la interpretación.

Significancia práctica. Consiste en aplicar una regla empírica a la matriz de factores. Si las cargas factoriales mayores a ± 0.30 se consideran que tienen un nivel mínimo; las cargas de ± 0.40 se consideran más importantes y las cargas de ± 0.50 o superiores se consideran prácticamente significativas. Así cuanto mayor sea el tamaño de la carga factorial, más importante resulta la carga al interpretar la matriz factorial. Dado que la carga factorial representa la correlación entre la variable observada y el factor, para que un factor explique un 50% de la varianza de la variable ha de tener una carga factorial mayor a 0.70 , lo anterior corresponde cuando se tiene un tamaño de muestra superior a 100 observaciones.

Significancia estadística. Al determinar el nivel de significación para la interpretación de las cargas, se debería emplear una aproximación similar a la utilizada para significación estadística de los coeficientes de correlación. Sin embargo, los errores estándar de las cargas factoriales son mayores que las correlaciones. Por lo tanto es necesario aplicar el concepto de potencia estadística para especificar la significancia estadística sujeta al tamaño de la muestra. La tabla

que se muestra a continuación contienen los tamaños muestrales necesarios que los valores de las cargas factoriales se consideren significativos.

Carga Factorial	Tamaño de la muestra necesario para la significancia ($\alpha=0.05$)
0.30	350
0.35	250
0.40	200
0.45	150
0.50	120
0.55	100
0.60	85
0.65	70
0.75	50

Ajuste basado en el número de variables. La desventaja de las consideraciones hechas anteriormente es que no se toma en cuenta el número de variables y los factores que se analizan. Según el número de variables analizadas, se incrementa el nivel de aceptable para considerar significativa una carga que decrece. El ajuste cobra importancia según se desplaza del primer factor extraído a los últimos.

En resumen se establecen las siguientes consideraciones:

- A mayor tamaño muestral, menor puede ser la carga para ser considerada significativa;
- A mayor número de variables analizadas, menor ha de ser la carga a ser considerada como significativa;
- A mayor número de factores, mayor ha de ser la carga de los últimos factores considerados para que la interpretación sea significativa.

8. Estimación de las puntuaciones de los factores

Hasta este momento se ha obtenido un modelo exploratorio de los datos mediante la técnica de AF para extraer información sobre variables latentes que no pueden ser medidas directamente y se la ha dado una interpretación a las misma. Corresponde ahora estimar los valores que corresponde a cada factor o variable latente identificado en cada uno de los individuos del estudio. Estos valores son útiles por:

- Representan un resumen parsimonioso de los datos originales útil para análisis subsecuentes.
- Ellos son una medición “pura” de las variables latentes, mientras que las variables observadas pueden ser ambiguas porque no se conoce de antemano que combinación de variables latentes puede estar representada por los valores observados.

El cálculo de las *puntuaciones (scores) factoriales* no es tan directo debido a que en la ecuación del modelo factorial las variables observadas están expresadas en termino de los factores. Por lo tanto, es necesario realizar un cambio en la relación. Esto es realizar la transformación del modelo de tal forma que

$$\hat{\mathbf{f}} = \hat{\mathbf{\Lambda}}^T \mathbf{S}^{-1} \mathbf{X}$$

Existen varios métodos para estimar las *puntuaciones factoriales*, algunos descritos por Rencher (1995) o resumidos por Hershberger (2005). Se podría decir que la principal dificultad con el Análisis de Factores no es la unicidad de la solución al tener la posibilidad de rotar la matriz de cargas, es la determinación de las puntuaciones factoriales.

9. Validación del modelo resultante

Es importante evaluar el grado de generalidad de los resultados para la población. La forma más directa de validación de los resultados consiste en adoptar una perspectiva de confirmación al valorar la *replicabilidad* de los resultados. Esto se realiza ya sea dividiendo la muestra o bien con una muestra adicional para repetir el análisis. También se aconseja realizar un *Análisis Factorial Confirmatorio*.

Es conveniente verificar la estabilidad de los resultados del modelo. Esta estabilidad depende de principalmente del tamaño de la muestra y el número de casos por variable. Si el tamaño de la muestra original lo permite puede dividirse aleatoriamente la muestra en dos subconjuntos, estimar el modelo factorial de cada subconjunto y comparar las dos matrices de cargas factoriales resultantes. Con el mecanismo anterior se está verificando la robustez de los factores encontrados.