



ANÁLISIS ESTADÍSTICO MULTIVARIADO

Análisis Multivariado

1 Análisis Factorial

- Introducción
- Estimación del modelo
- Diagnóstico del modelo
- Rotación y estimación de factores

2 ANEXO: Rotación Varimax

Modelo de análisis factorial

- El análisis factorial tiene por objetivo explicar un conjunto de variables observadas por un pequeño número de variables latentes o no observadas que llamaremos factores.
- Sea x un vector de p variables aleatorias referidas a una población de interés. El modelo de análisis factorial establece la siguiente relación:

$$x = \mu + \Lambda f + u$$

- $f_{(m \times 1)}$ es un vector de variables latentes o factores no observados. Supondremos que $f \sim N_m(0, I)$.
- $\Lambda_{(p \times m)}$ es una matriz de constantes desconocidas ($m < p$) denominada matriz de cargas.
- $u_{(p \times 1)}$ es un vector de errores no observados. Recoge el efecto de todas las variables ajenas a los factores que influyen sobre x . Supondremos que $u \sim N_p(0, \Psi)$, donde Ψ es diagonal y los errores están no correlacionados con los factores f .

Modelo de análisis factorial

Del modelo se deduce:

- La media de las variables x es $E(x) = \mu$; tanto los factores f como los errores u tienen media cero.
- $x \sim N_p(\mu, \Sigma)$.
- La ecuación que define al modelo factorial implica que dada una muestra aleatoria de n elementos generada por el modelo factorial, cada dato x_{ij} puede escribirse como:

$$x_{ij} = \mu_j + \lambda_{j1}f_{1i} + \lambda_{j2}f_{2i} + \cdots + \lambda_{jm}f_{mi} + u_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, p$$

que descompone el valor observado en el individuo i de la variable j como la suma de $m + 2$ términos.

Modelo de análisis factorial

- Agrupando la información para todas las variables y observaciones podemos escribir la matriz X de datos como:

$$X = 1\mu' + F\Lambda' + U$$

- Donde 1 es un vector $(n \times 1)$ de unos, F es una matriz $(n \times m)$ que contiene los valores correspondientes a los m factores para los n elementos de la población, Λ' es la matriz $(m \times p)$ de cargas factoriales cuyos coeficientes relacionan las variables y los factores y finalmente U es una matriz $(n \times p)$ de errores.
- La matriz de covariancias entre las variables y los factores se obtienen como:

$$E[(x - \mu)f'] = E[(\Lambda f + u)f'] = \Lambda E[ff'] + E[uf'] = \Lambda$$

- Recordar que por hipótesis los factores están no correlacionados, tienen media cero y no están correlacionados con los errores.

Modelo de análisis factorial

- Los términos λ_{ij} representan la covariancia entre la variable x_i y el factor f_j . Como los factores tienen variancia unitaria estos son los coeficientes de regresión cuando explicamos las variables observadas por medio de los factores.
- La matriz de variancias y covariancias de las observaciones verifica:

$$\begin{aligned}\Sigma &= E[(x - \mu)(x - \mu)'] = E[(\Lambda f + u)(\Lambda f + u)'] \\ &= \Lambda E[ff']\Lambda' + E[uu'] = \Lambda\Lambda' + \Psi\end{aligned}$$

- La matriz de variancias y covariancias de las observaciones admite una descomposición como suma de dos matrices.

Modelo de análisis factorial

- La matriz $\Lambda\Lambda'$ es simétrica de rango $m < p$. Representa la variancia común al conjunto de las p variables y depende de las covariancias entre las variables y los factores.
- La segunda matriz, Ψ , es diagonal y representa la variancia que es específica de cada variable e independiente del resto.
- Esta descomposición implica que las variancias de las variables observadas pueden expresarse como:

$$\sigma_i^2 = \sum_{j=1}^m \lambda_{ij}^2 + \psi_i^2 = h_i^2 + \psi_i^2$$

- El primer término es la suma de los efectos de los factores y el segundo término es el efecto del error específico.

Modelo de análisis factorial

- Al primer término h_i^2 lo llamaremos **comunalidad**. La expresión anterior puede interpretarse como una descomposición de la variancia

Variancia observada = variabilidad común + variabilidad específica
que es análoga a la descomposición clásica de la variabilidad de los datos en una parte de variancia explicada y otra no explicada.

- En el modelo factorial la parte de variancia explicada es debida a los factores y la no explicada se debe al "ruido" o componente aleatorio específico.
- En el modelo factorial ni la matriz de cargas Λ ni los factores f son observables.
- Esto plantea un problema de indeterminación: dos representaciones (Λ, f) y (Λ^*, f^*) serán equivalentes si $\Lambda f = \Lambda^* f^*$. Esta situación conduce a dos tipos de indeterminaciones:
 - ▶ Un conjunto de datos puede explicarse con la misma precisión con factores correlacionados o no correlacionados.
 - ▶ Los factores no quedan determinados de manera única.

Unicidad del modelo

Un conjunto de datos puede explicarse con la misma precisión con factores correlacionados o no correlacionados.

- Si H es cualquier matriz no singular, el modelo factorial $x = \mu + \Lambda f + u$ puede escribirse como $x = \mu + \Lambda H H^{-1} f + u$ y llamando $\Lambda^* = \Lambda H$ a la nueva matriz de cargas y $f^* = H^{-1} f$ a los nuevos factores se obtiene $x = \mu + \Lambda^* f^* + u$.
- Los nuevos factores tienen distribución $f^* \sim N(0, H^{-1}(H^{-1})')$ y por lo tanto están correlacionados.
- De manera análoga se demuestra que podemos partir de un modelo con factores correlacionados y encontrar una expresión equivalente de las variables a partir de factores no correlacionados.

Unicidad del modelo

Los factores no quedan determinados de manera única.

- En efecto, si H es ortogonal, las representaciones $x = \mu + \Lambda f + u$ y $x = \mu + \Lambda H H' f + u$ son indistinguibles.
- Ambas representaciones están expresadas en función de factores no correlacionados, cuya matriz de variancias y covariancias es la matriz identidad.
- El modelo factorial no está identificado ante rotaciones ortogonales de los factores.
- Esta indeterminación se resuelve imponiendo restricciones sobre los componentes de la matriz de cargas factoriales, como veremos a continuación. Consideraremos dos restricciones alternativas.

Unicidad del modelo

Primer Criterio: imponer $\Lambda'\Lambda = D = \text{Diagonal}$

- Con esta normalización los vectores que definen el efecto de cada factor sobre las p variables observadas son ortogonales. Los factores además de estar no correlacionados producen efectos lo mas distintos posible en las variables.
- Se puede demostrar que esta normalización define un única matriz de cargas factoriales. Además, cuando se verifica esta normalización podemos ver que las columnas de la matriz Λ son los autovectores asociados a la matriz $(\Sigma - \Psi)$, cuyos autovalores constituyen los elementos diagonales de la matriz D .
- Este principio se utiliza para la estimación del modelo factorial mediante el método del factor principal.

Unicidad del modelo

Segundo Criterio: imponer $\Lambda' \Psi^{-1} \Lambda = D = \text{Diagonal}$

- Con esta normalización los vectores que definen el efecto de cada factor sobre las p variables, ponderados por las variancias de los errores de cada ecuación, están no correlacionados.
- Nuevamente esta normalización define una única matriz de cargas factoriales. Cuando se verifica esta normalización podemos ver que la matriz $\Psi^{-1/2} \Lambda \Psi^{-1/2}$ tiene autovectores $\Psi^{-1/2} \Lambda$ cuyos autovalores corresponden a los elementos diagonales de la matriz $(D + I)$.
- Este principio se utiliza para la estimación del modelo factorial mediante el método de máxima verosimilitud.

Estimación: método del factor principal

- Este método está basado en componentes principales y se utiliza para estimar la matriz de cargas factoriales.
- La estimación se plantea a partir del principio de mínimos cuadrados donde el objetivo es minimizar la suma de cuadrados de los elementos $S - \Sigma = S - \Lambda\Lambda' - \Psi$

$$F_{LS}(\Lambda, \Psi) = \text{tr}[(S - \Sigma)'(S - \Sigma)] = \text{tr}(S - \Sigma)^2 = \text{tr}(S - \Lambda\Lambda' - \Psi)^2$$

- La minimización de la expresión anterior conduce a las siguientes ecuaciones normales:

$$(S - \hat{\Lambda}\hat{\Lambda}' - \hat{\Psi})\hat{\Lambda} = 0$$

$$\text{diag}(S - \hat{\Lambda}\hat{\Lambda}' - \hat{\Psi}) = 0$$

Estas ecuaciones no se resuelven directamente, en general se toma como punto de partida una estimación inicial de la matriz de variancias específicas.

Estimación: método del factor principal

- Supongamos que podemos obtener una estimación inicial de la matriz de variancias de los errores $\hat{\Psi}$. Entonces podemos calcular:

$$S - \hat{\Psi} = \Lambda \Lambda'$$

es una matriz simétrica se puede descomponer:

$$S - \hat{\Psi} = HGH' = (HG^{1/2})(HG^{1/2})'$$

donde H es cuadrada de orden p y ortogonal, G también es de orden p , diagonal y contiene los autovalores de $S - \hat{\Psi}$. El modelo factorial establece que G tiene la siguiente forma:

$$G = \begin{pmatrix} G_{1(m \times m)} & 0_{(m \times (p-m))} \\ 0_{((p-m) \times m)} & 0_{((p-m) \times (p-m))} \end{pmatrix}$$

ya que $S - \hat{\Psi}$ tiene rango m .

Estimación: método del factor principal

- Si llamamos H_1 a la matriz $(p \times m)$ que contiene los autovectores asociados a los autovalores de G_1 podemos tomar como estimador de Λ la matriz $(p \times m)$ $\hat{\Lambda} = H_1 G_1^{1/2}$ Con lo cual resolvemos el problema. Observemos que la normalización resultante es:

$$\hat{\Lambda}'\hat{\Lambda} = (G_1^{1/2} H_1')(H_1 G_1^{1/2})$$

- Ya que los autovectores de matrices simétricas son ortogonales, por lo que $H_1' H_1 = I_m$.
- Con este método se obtienen estimadores de Λ con columnas ortogonales entre si.
- En la práctica este procedimiento se lleva a cabo de manera iterativa, veremos a continuación los pasos a seguir...

Estimación: método del factor principal

- 1 Partimos de una estimación inicial de Ψ
- 2 Calculamos la matriz cuadrada y simétrica $Q_i = S - \hat{\Psi}$
- 3 Obtenemos la descomposición espectral de Q_i como:

$$Q_i = H_{1i} G_{1i} H'_{1i} + H_{2i} G_{2i} H'_{2i}$$

donde G_{1i} contiene los m autovalores mas grandes de Q_i y H_{1i} sus respectivos autovectores.

- ▶ Elegiremos m de manera que los restantes autovalores, contenidos en G_{2i} sean pequeños y de tamaño similar.
 - ▶ La matriz Q_i puede no ser definida positiva. Esto no es un problema grave si los autovalores negativos son pequeños y podemos suponerlos próximos a 0.
- 4 Tomar $\hat{\Lambda}_{i+1} = H_{1i} G_{1i}^{1/2}$ y volver al paso 1. Iterar hasta que haya convergencia, es decir hasta que $|\hat{\Lambda}_{i+1} - \hat{\Lambda}_i| < \varepsilon$.

Estimación: método del factor principal

- Como resultado de esta estimación se obtienen estimadores consistentes pero no eficientes, como en el caso de máxima verosimilitud.
- No serán invariantes ante transformaciones lineales: no necesariamente obtendremos los mismos resultados si partimos de la matriz de variancias y covariancias o de la matriz de correlaciones.
- Para implementar este procedimiento en la práctica debemos especificar como obtenemos la estimación inicial $\hat{\Psi}_i$.
- Estimar los términos diagonales de la matriz Ψ equivale a definir valores para los términos diagonales h_i^2 de $\Lambda\Lambda'$ ya que $h_i^2 = s_i^2 - \Psi_{ii}$.

Estimación: método del factor principal

- Tomar $\hat{\Psi}_{ii} = 0$, esto equivale a extraer los componentes principales de S .
- Tomar $\hat{\Psi}_{ii} = 1/s_{ii}^*$ donde s_{ii}^* es el elemento diagonal i -ésimo de la matriz de precisión S^{-1} . Se puede demostrar que esto equivale a tomar $\hat{h}_i^2 = s_i^2 - s_i^2(1 - R_i^2) = s_i^2 R_i^2$.
- 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.

Estimación: máxima verosimilitud

- Bajo el modelo factorial la función de densidad de las observaciones originales es $N_p(\mu, \Sigma)$. Por lo tanto la verosimilitud es la que ya analizamos.
- Sustituyendo μ por su estimador \bar{x} , la función soporte para Σ es la siguiente:

$$\text{Log}(\Sigma/X) = -\frac{n}{2} \log|\Sigma| - \frac{n}{2} \text{tr}(S\Sigma^{-1})$$

y sustituyendo Σ en función de Λ y ψ obtenemos:

$$L(\Lambda, \psi) = -\frac{n}{2} \log|\Lambda\Lambda' + \Psi| - \frac{n}{2} \text{tr}(S(\Lambda\Lambda' + \Psi)^{-1})$$

Estimación: máxima verosimilitud

- Los estimadores de máxima verosimilitud se obtienen maximizando la expresión anterior con respecto a las matrices Λ y Ψ .
- Derivando con respecto a estas matrices y operando algebraicamente se obtienen las siguientes ecuaciones normales:

$$\hat{\Psi} = \text{diag}(S - \hat{\Lambda}\hat{\Lambda}')$$

$$(\hat{\Psi}^{-1/2}S\hat{\Psi}^{-1/2} - I)(\hat{\Psi}^{-1/2}\hat{\Lambda}) = (\hat{\Psi}^{-1/2}\hat{\Lambda})D$$

donde D es la matriz que satisface la normalización que vimos antes:

$$(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda}) = D = \text{Diagonal}$$

Estimación: máxima verosimilitud

- Con estas tres ecuaciones se puede resolver el sistema utilizando un algoritmo iterativo, por ejemplo Newton-Raphson.
- La solución numérica a veces resulta difícil porque es posible que $\hat{\Psi}$ no sea definida positiva.
- Cuando se obtienen estimaciones ψ que son nulas o negativas se obtiene lo que se conoce como *Heywood case*. Cuando esto ocurre es una indicación de que el modelo factorial no es una representación adecuada de los datos.
- Observemos que la segunda ecuación implica una ecuación de autovalores: $(\hat{\Psi}^{-1/2}\hat{\Lambda})$ contiene los autovectores asociados a la matriz simétrica $(\hat{\Psi}^{-1/2}S\hat{\Psi}^{-1/2} - I)$ y D contiene sus autovalores.

Estimación: máxima verosimilitud

El algoritmo iterativo para resolver estas ecuaciones es el siguiente:

- Comenzamos con una estimación inicial. Si tenemos una estimación $\hat{\Lambda}_i$, con $i = 1$ la primera vez, por ejemplo por el método del factor principal, se calcula la matriz $\hat{\Psi}_i$ mediante $\hat{\Psi}_i = \text{diag}(S - \hat{\Lambda}_i \hat{\Lambda}_i')$.
- Se calcula la matriz cuadrada y simétrica A_i :

$$A_i = (\hat{\Psi}_i^{-1/2}(S - \hat{\Psi}_i)\hat{\Psi}_i^{-1/2}) = \hat{\Psi}_i^{-1/2}S\hat{\Psi}_i^{-1/2} - I$$

Esta matriz pondera los términos de S por su importancia en términos de los componentes específicos.

Estimación: máxima verosimilitud

- Se obtiene la descomposición espectral de A_i :

$$A_i = H_{1i} G_{1i} H'_{1i} + H_{2i} G_{2i} H'_{2i}$$

donde los m autovalores mas grandes de A_i están en la matriz diagonal G_{1i} de dimensión $(m \times m)$ y los $p - m$ más chicos en la diagonal de G_{2i} mientras que las matrices H_{1i} y H_{2i} contienen los respectivos autovectores.

- Se calcula: $\hat{\Lambda}_{i+1} = \hat{\Psi}_i^{1/2} H_{1i} G_{1i}^{1/2}$ y se sustituye en la función de verosimilitud, que se maximiza con respecto a Ψ .
- Con el resultado obtenido se vuelve al paso 2 hasta lograr la convergencia del procedimiento.
- El resultado de la estimación máximo verosímil no depende del uso de S o R .

Determinación del número de factores

- Supongamos que se ha estimado un modelo con m factores. El test para comprobar si la descomposición es adecuada puede plantearse como un test de razón de verosimilitud:

$$H_0) \Sigma = \Lambda\Lambda' + \Psi \qquad H_1) \Sigma \neq \Lambda\Lambda' + \Psi$$

El test estadístico que debemos calcular es $\lambda = 2(\ln(H_1) - \ln(H_0))$. Llamemos como $\hat{\Sigma}_0$ el valor de la matriz de variancias y covariancias de los datos estimados bajo H_0 . Las funciones de log-verosimilitud evaluadas en cada una de las hipótesis son:

$$\ln(H_0) = -\frac{n}{2} \log|\hat{\Sigma}_0| - \frac{n}{2} \text{tr}(S\hat{\Sigma}_0^{-1})$$

$$\ln(H_1) = -\frac{n}{2} \log|S| - \frac{np}{2}$$

Determinación del número de factores

- Por lo tanto el estadístico del test de razón de verosimilitudes es:

$$\lambda = n(\log|\hat{\Sigma}_0| + \text{tr}(S\hat{\Sigma}_0^{-1}) - \log|S| - p)$$

- El estimador MV de la matriz Σ bajo $H_0, \hat{\Sigma}_0$, minimiza la distancia a S medida con la traza, es decir

$$\text{tr}(S\hat{\Sigma}_0^{-1}) = p$$

En consecuencia, la expresión del estadístico se reduce a

$$\lambda = n \log \left(\frac{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|}{|S|} \right)$$

que mide la distancia entre $\hat{\Sigma}_0$ y S en términos del determinante.

- El test conduce al rechazo de H_0 cuando λ supere al valor crítico asociando a la distribución χ^2 con g grados de libertad y al nivel de significación elegido. El valor de $g = ((p - m)^2 - (p + m))/2$

Determinación del número de factores

- Bartlett (1954) ha demostrado que la aproximación asintótica de la distribución χ^2 mejora en muestras finitas introduciendo un factor de corrección en reemplazo de n . Con esta corrección se rechaza H_0 si:

$$\left(n - 1 - \frac{2p + 4m + 5}{6} \right) \log \left(\frac{|\widehat{\Lambda}\widehat{\Lambda}' + \widehat{\Psi}|}{|S|} \right) > \chi_{g,\alpha}^2$$

Generalmente este test se aplica secuencialmente: se estima el modelo con $m = m_1$ factores y se prueba H_0 . Si se rechaza, se reestima el modelo con $m = m_1 + 1$ factores y se prueba H_0 , hasta encontrar el valor de m para el cual no se rechaza la hipótesis nula.

Diagnóstico del modelo

- Para verificar si el modelo es adecuado conviene calcular los factores y los residuos para analizar sus propiedades.
- De acuerdo con las hipótesis del modelo, $u \sim N_p(0, \Psi)$, por lo tanto si la matriz de variancias y covariancias de los residuos no es diagonal deberíamos aumentar el número de factores hasta que los residuos estimados verifiquen esa hipótesis. Por otro lado es posible calcular estadísticos de bondad del ajuste:
- Coeficiente de correlación al cuadrado entre cada variable observable y los factores:

$$\gamma_i^2 = h_i^2 / s_i^2 = 1 - \psi_i^2 / s_i^2$$

- Coeficiente de determinación:

$$R^2 = 1 - \left(\frac{|\hat{\Psi}|}{|\hat{\Sigma}|} \right)^{1/p}$$

Diagnóstico del modelo

- Otro criterio de análisis es calcular el Error Cuadrático Medio Residual:

$$RMS = \sqrt{\frac{tr(S - \hat{\Lambda}\hat{\Lambda}' - \hat{\Psi})^2}{p(p-1)/2}}$$

- Si existe una estructura factorial en los datos (*strong data*), los distintos estadísticos de bondad de ajuste deberían indicar aproximadamente el mismo número de factores.
- *Strong data* se refiere a que las variables deben presentar un grado de asociación (no pueden ser independientes).

Diagnóstico del modelo

- Una medida para determinar si tenemos *strong data* es la siguiente:

$$MSA = \frac{\sum_{i < j} r_{ij}^2}{\sum_{i < j} r_{ij}^2 + \sum_{i < j} q_{ij}^2}$$

donde $R = (r_{ij})$ y $Q = DR^{-1}D = (q_{ij})$, $D = ((\text{Diag}(R^{-1}))^{1/2})^{-1}$.

- Si $R \rightarrow I$, el criterio MSA tiende a cero.
- Para evaluar la asociación entre las variables de la matriz de datos a los efectos de aplicar el análisis factorial, Kaiser y Rice (1974) recomiendan que el estadístico MSA sea mayor o igual a 0.8.

Rotación de factores

- La matriz de cargas no está identificada ante rotaciones ortogonales de los factores.
- Está definido el espacio de las columnas de la matriz de cargas pero cualquier base de ese espacio puede ser una solución.
- La elección se realiza tomando en cuenta la interpretación de los factores.
- Intuitivamente será más fácil interpretar un factor cuando se asocia a un grupo de variables observadas y no a todas ellas.
- Existen diversos criterios para definir la rotación de factores.

Rotación de factores

- Thurstone (1947) definió la idea de estructura simple:
- Cada fila de la matriz Λ debería tener al menos un valor cero.
- Cada columna de la matriz Λ debería tener al menos m valores iguales a cero.
- Para todos los pares de columnas de la matriz Λ debería haber varias filas en las que una carga es cero y otra es distinta de cero, y solo unas pocas filas debieran tener ambas cargas distintas de cero.
- Si $m \geq 4$, varios pares de columnas de Λ deberían tener dos cargas iguales a 0.
- Estas condiciones son generales y no se verifican estrictamente.

Criterios de rotación ortogonal

- **Criterio Varimax:** El criterio que se utiliza para definir esta rotación es el de maximizar la variancia de los coeficientes que definen los efectos de cada factor sobre las variables observadas.
- **Criterio Quartimax:** maximiza la variancia a través de los factores (filas de la matriz Λ).
- Este procedimiento generalmente da por resultado un único factor con cargas altas y otros factores con cargas diferentes a través de las variables.
- **Criterio Equamax:** es una transformación ortogonal que maximiza la suma ponderada de las variancias entre filas y entre columnas de la matriz de cargas.
- Es un caso intermedio entre Varimax y Quartimax.

Rotación Oblicua

- El modelo factorial también está indeterminado ante rotaciones oblicuas.
- A partir de una estimación inicial de los factores (f , no correlacionados entre si) es posible definir nuevos factores $f^* = Hf$ siendo H una matriz no singular.
- La nueva matriz de variancias y covariancias de los factores será $\Sigma_f^* = HH'$.
- Los nuevos factores así obtenidos no pueden interpretarse de manera independiente.

Estimación de los factores

- Los dos enfoques que se pueden encontrar en la literatura son los siguientes:
 - 1 Suponer que los factores para cada observación son parámetros a estimar (Método de Bartlett).
 - 2 Suponer que los valores de los factores son variables aleatorias.
- En el primer caso, para el individuo i , $x_i \sim N_p(\Lambda f_i, \Psi)$. Los parámetros f_i pueden estimarse por máxima verosimilitud, y el estimador resultante coincide con el que se obtiene por mínimos cuadrados generalizados:

$$\hat{f}_i = (\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} x_i$$

- Si conocemos Λ , el modelo factorial $x_i = \Lambda f_i + u_i$ es un modelo de regresión con variable dependiente x_i , donde las variables explicativas son las columnas de Λ y parámetros desconocidos f_i . Como el error $u_i \sim N(0, \Psi)$ se utiliza mínimos cuadrados generalizados que da por resultado el estimador de los factores f_i presentados arriba.

Estimación de los factores

- Sea x_i el vector $(p \times 1)$ en el individuo i . Su función de densidad será:

$$f(x_i) = |\Psi|^{-1/2} (2\pi)^{-p/2} \exp\{-1/2(x_i - \Lambda f_i)' \Psi^{-1} (x_i - \Lambda f_i)\}$$

supongamos que Ψ y Λ son conocidas y se trata de estimar f_i . Entonces la función de log-verosimilitud será:

$$L = \log f(x_i) = K - 1/2(x_i - \Lambda f_i)' \Psi^{-1} (x_i - \Lambda f_i)$$

donde K es una constante.

- Maximizar L conduce al criterio de mínimos cuadrados.

$$M = (x_i - \Lambda f_i)' \Psi^{-1} (x_i - \Lambda f_i)$$

Estimación de los factores

- Entonces,

$$M = x_i' \Psi^{-1} x_i - 2f_i' \Lambda' \Psi^{-1} x_i + f_i' \Lambda' \Psi^{-1} \Lambda f_i$$

- Derivando con respecto a f_i e igualando a cero

$$\frac{\partial M}{\partial f_i} = 0 = -2\Lambda' \Psi^{-1} x_i + 2\Lambda' \Psi^{-1} \Lambda f_i$$

por lo tanto,

$$f_i = (\Lambda' \Psi^{-1} \Lambda)^{-1} \Lambda' \Psi^{-1} x_i$$

- Sustituyendo en esta expresión Λ y ψ por sus estimadores máximo verosímiles se obtiene el vector \hat{f}_i para cada observación.

Estimación de los factores

- Segundo caso: suponer que los valores de los factores son variables aleatorias.
- Este procedimiento implica buscar un predictor lineal que minimice el error cuadrático medio de la predicción. Llamando f_i a los valores de los factores para el individuo i y x_i representa al vector de variables observadas, el vector (f_i, x_i) tendrá distribución normal multivariada y el objetivo es encontrar $E[f_i/x_i]$.

$$E[f_i/x_i] = E[f_i] + \text{cov}(f_i, x_i) \text{var}(x_i)^{-1} (x_i - E[x_i])$$

Como $E[f_i] = 0$ y $\text{Cov}(f_j, x_k) = \lambda_{kj}$, asumiendo que las variables observadas tienen media cero, podemos escribir:

$$\hat{f}_i = \widehat{E[f_i/x_i]} = \hat{\Lambda}' \hat{\Sigma}^{-1} x_i$$

Estimación de los factores

- Utilizando la siguiente propiedad:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$$

resulta:

$$\hat{\Sigma}^{-1} = (\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})^{-1} = \hat{\Psi}^{-1} - \hat{\Psi}^{-1}\hat{\Lambda}(I + \hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}$$

Estimación de los factores

- Entonces:

$$\widehat{\Lambda}'\widehat{\Sigma}^{-1} = \widehat{\Lambda}'\widehat{\Psi}^{-1} - \widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda}(I + \widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda})^{-1}\widehat{\Lambda}'\widehat{\Psi}^{-1}$$

$$\widehat{\Lambda}'\widehat{\Sigma}^{-1} = [I - \widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda}(I + \widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda})^{-1}]\widehat{\Lambda}'\widehat{\Psi}^{-1}$$

$$\widehat{\Lambda}'\widehat{\Sigma}^{-1} = (I + \widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda})^{-1}\widehat{\Lambda}'\widehat{\Psi}^{-1}$$

y sustituyendo en la expresión de \hat{f}_i se obtiene:

$$\hat{f}_i = (I + \widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda})^{-1}\widehat{\Lambda}'\widehat{\Psi}^{-1}x_i$$

- Este estimador de los factores es el predictor lineal de los factores sobre los datos.

Análisis factorial y componentes principales

- En componentes principales descomponemos la matriz de variancias y covariancias de X como $S = ADA'$. La matriz A contiene los autovectores asociados a la matriz S mientras que D contiene en la diagonal principal los autovalores de S .
- Si $\lambda_j = 0$ para $j < p$ podemos reconstruir S a partir de los primeros j componentes principales. Llamando $H = AD^{1/2}$ tenemos que $S = HH'$.
- En análisis factorial descomponemos S como $S = \Lambda\Lambda' + \Psi$ y como Ψ es diagonal puede recoger las variancias de las variables, mientras que la matriz de cargas recoge las covariancias entre las variables.
- Esta es la diferencia más importante entre ambos métodos: componentes principales trata de explicar las variancias mientras que el análisis factorial explica las covariancias o correlaciones entre las variables.

Análisis factorial y componentes principales

- Observemos que si $\Psi \approx 0$ es lo mismo tomar m componentes principales que estimar m factores.
- La diferencia es tanto menor cuanto menor sea $\hat{\Psi}$.
- Otro enfoque para analizar la relación entre ambos métodos: sea X la matriz de datos originales y Z la matriz con los valores de las componentes principales. Entonces $Z = XA$ y como A es ortogonal también podemos expresar $X = ZA'$, lo que nos permite recuperar las variables originales a partir de los componentes:

$$x_j = \alpha_{j1}z_1 + \cdots + \alpha_{jm}z_m + \cdots + \alpha_{jp}z_p \quad (j = 1, \dots, p)$$

- Si solo utilizamos m componentes podemos escribir:

$$x_j = \alpha_{j1}z_1 + \cdots + \alpha_{jm}z_m + v_j \quad (j = 1, \dots, p)$$

Análisis factorial y componentes principales

- Esta última representación es análoga al modelo factorial, ya que v_j no estará correlacionada con los factores $(\alpha_{j1}z_1 + \dots + \alpha_{jm}z_m)$ al incluir únicamente las variables (z_{m+1}, \dots, z_p) que no están correlacionados con los primeros m factores.
- En el modelo factorial los errores de las distintas ecuaciones no están correlacionados mientras que en esta representación lo estarán.
- Estos resultados indican que si existen m componentes principales que explican una proporción muy alta de la variabilidad total, de manera que la variabilidad específica dada por los términos diagonales de sea pequeña, el análisis factorial y el análisis de componentes principales sobre la matriz de correlaciones darán resultados similares.
- En este caso, la estimación mediante el método del factor principal dará resultados similares al de máxima verosimilitud.

Rotación Varimax

- El criterio que se utiliza para definir esta rotación es el de maximizar la variancia de los coeficientes que definen los efectos de cada factor sobre las variables observadas.
- Llamaremos δ_{ij} a los coeficientes de la matriz de carga asociado al factor j en las i ecuaciones después de la rotación.
- Llamaremos δ_j al vector correspondiente a la columna j de la matriz de cargas luego de la rotación.
- Se desea que la variancia de los coeficientes (elevados al cuadrado para prescindir del signo) del vector sea máxima.

Rotación Varimax

- Definimos el valor promedio:

$$\bar{\delta}_{.j} = \frac{1}{p} \sum_{i=1}^p \delta_{ij}^2$$

- La variabilidad para el factor j se define como:

$$\frac{1}{p} \sum_{i=1}^p (\delta_{ij}^2 - \bar{\delta}_{.j})^2 = \frac{1}{p} \sum_{i=1}^p \delta_{ij}^4 - \frac{1}{p^2} \left(\sum_{i=1}^p \delta_{ij}^2 \right)^2$$

- El criterio *varimax* maximiza la suma de las variancias para todos los factores:

$$q = \frac{1}{p} \sum_{j=1}^m \sum_{i=1}^p \delta_{ij}^4 - \frac{1}{p^2} \left(\sum_{j=1}^m \sum_{i=1}^p \delta_{ij}^2 \right)^2$$

Rotación Varimax

- Sea Λ la matriz de cargas estimada originalmente. El problema es hallar una matriz ortogonal M tal que $\delta = \Lambda M$ cuyos coeficientes verifiquen la condición de maximizar q .
- Definimos los elementos de δ como $\delta_{ij} = \lambda'_i m_j$, donde λ'_i es la fila i de la matriz Λ y m_j es la columna j de la matriz M .
- Los elementos de la matriz M se obtienen derivando q con respecto a cada uno de los términos m_{ij} teniendo en cuenta las restricciones de ortogonalidad $m'_k m_k = 1$, $m'_k m_h = 0$ para $k \neq h$.