

Análisis de factores

Nuevamente el vector $x' = (x_1, \dots, x_p)$ contiene las p variables que interesan estudiar los n individuos de la muestra. Las variables consideradas tienen una escala de medición **continua**.

La correlación r_{ij} entre dos variables x_i y x_j puede deberse a que ambas tienen una relación con otra variable z_k que se considera como no observable, es decir que si x_i está relacionada con z_k y x_j está relacionada con z_k , entonces debe haber una relación entre x_i y x_j .

El coeficiente de correlación parcial $r_{ij.k}$ mide la asociación que hay entre x_i y x_j luego de remover el efecto que tiene z_k en cada una. Se puede pensar que si el valor de $r_{ij.k}$ se acerca a cero, sucede que z_k ha explicado la correlación existente entre x_i y x_j ; pueden considerarse varias variables z_1, \dots, z_m para explicar la asociación entre x_i y x_j .

La idea en análisis de factores es explicar las correlaciones contenidas en $R = [r_{ij}]$ a través de un conjunto de variables no observables f_1, \dots, f_m de manera que las **correlaciones parciales** $r_{ij.f_1, \dots, f_m}$ sean muy pequeñas, se busca también que esa m sea pequeña.

Por ejemplo las calificaciones que los estudiantes puedan sacar en cálculo mental, manejo de vocabulario, comprensión de lectura, dibujo, etc, están correlacionadas entre sí, ya que cada estudiante posee cualidades como inteligencia, memoria, habilidad espacial, etc, todas estas cualidades no pueden medirse directamente por lo que se les considera como **no observables**.

0.1. Modelo Básico

En análisis de factores se quiere explicar las relaciones entre las variables manifiestas x_1, \dots, x_p a través de m variables latentes f_1, \dots, f_m que en el modelo se llaman factores comunes, de la siguiente forma:

$$x_j = \mu_j + \lambda_{j1}f_1 + \dots + \lambda_{jm}f_m + e_j$$

Las μ_j y las λ_{ji} son constantes, mientras que e_j y los f_i (con $j = 1 \dots p$ y $i = 1 \dots m$) son variables aleatorias. Se considera a los e_j **no** correlacionados entre sí y los e_j son **no** correlacionados con f_i . A los e_j se les llama factores específicos.

De manera matricial se puede escribir como

$x = \mu + \Lambda f + e$ o centrando la variable en cero $x - \mu = \Lambda f + e$ donde $x = (x_1, x_2, \dots, x_p)$, $e' = [e_1, e_2, \dots, e_p]$, $f' = [f_1, f_2, \dots, f_m]$ y la matriz Λ contiene sólo constantes

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \vdots & & & \vdots \\ \lambda_{p1} & \dots & \dots & \lambda_{pm} \end{bmatrix}$$

Se supone que e y f tienen media cero, esto es $E(e) = 0$ y $E(f) = 0$.

La matriz de varianzas y covarianzas del vector de factores específicos

$cov(e) = \Psi$ es diagonal

$$\Psi = \begin{bmatrix} \psi_1^2 & 0 & \dots & 0 \\ 0 & \psi_2^2 & \dots & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \psi_p^2 \end{bmatrix}$$

Suponiendo que la matriz de varianzas y covarianzas de f es Γ , por ser simétrica y semipositiva definida se puede escribir como $\Gamma = LL'$, si se considera el vector $f^* = L^{-1}f$, este vector mantiene su media en cero y tiene matriz de varianzas y covarianzas $var(f^*) = L^{-1}\Gamma L^{-1'} = L^{-1}LL'L^{-1'} = I$.

Ahora escribiendo $\Lambda^* = \Lambda L$ entonces $\Lambda^* f^* = \Lambda L L^{-1} f = \Lambda f$, haciendo esto que el modelo

$$x = \mu + \Lambda^* f^* + e = \mu + \Lambda L^{-1} f + e$$

resulte indistinguible del original, entonces sin pérdida de generalidad se consideran a los factores comunes como **no** correlacionados entre sí.

Como se han considerado los factores comunes con correlación cero al calcular la varianza de x_j se tiene que:

$$var(x_j) = \lambda_{j1}^2 + \lambda_{j2}^2 + \dots + \lambda_{jm}^2 + var(e_j) = \underbrace{\sum \lambda_{jk}^2}_{\text{comunalidad}} + \underbrace{\psi_j^2}_{\text{especificidad}}$$

Y las covarianzas quedan como:

$$cov(x_i, x_j) = \sum_{k=1}^m \lambda_{ik} \lambda_{jk}$$

Entonces el modelo de análisis de factores también permite expresar a la matriz de varianzas y covarianzas de x , la Σ como:

$$\Sigma = \Lambda \Lambda' + \Psi$$

Al calcular la covarianza entre los factores comunes y las variables originales se tiene que:

$$cov(x, f) = E((\Lambda f + e)f') = \Lambda$$

esto porque $E(ff') = I$ y $E(ef') = 0$, de manera que las cargas λ_{ij} son la covarianza entre la variable original x_i y el factor f_j .

Entonces el análisis de factores es adecuado para explicar las covarianzas y el análisis de componentes principales es adecuado para explicar las varianzas.

Los factores se calculan por tres métodos:

1. Extracción inicial de factores: factores principales
2. factores principales iterados

3. factores de máxima verosimilitud

0.1.1. Factores principales

$$y_1 = \alpha_{11}x_1 + \dots + \alpha_{1p}x_p$$

⋮

$$y_p = \alpha_{p1}x_1 + \dots + \alpha_{pp}x_p$$

como los eigenvectores α_j son ortogonales se tiene que

$$x_1 = \alpha_{11}y_1 + \dots + \alpha_{1p}y_p$$

⋮

$$x_p = \alpha_{p1}y_1 + \dots + \alpha_{pp}y_p$$

de tal forma que en los últimos términos se explica poco pues las componentes principales van reduciendo su varianza.

Podemos escribir lo anterior como

$$x_i = \alpha_{1i}y_1 + \alpha_{2i}y_2 + \dots + \alpha_{mi}y_m + \eta_i$$

que se parece a

$$x_i = \lambda_{1i}f_1 + \lambda_{2i}f_2 + \dots + \lambda_{mi}f_m + e_i$$

donde $\eta_i = \alpha_{m+1}y_{m+1} + \dots + \alpha_{py_p}$.

Hay que fijarse que los η_i no cumplen con el supuesto de ser no correlacionados. Como $\eta_i = \alpha_{m+1}y_{m+1} + \dots + \alpha_{py_p}$ para cada i aparecen las mismas componentes principales y_j , entonces las η_i sí están correlacionadas.

Sin embargo esto se toma como una aproximación.

Para asegurar que los factores comunes f_j tengan varianza 1 se redefinen como

$$f_j = \frac{y_j}{\sqrt{\text{var}(y_j)}}$$

Tomando esto en cuenta cada x_i queda definida como:

$$x_i = \alpha_{1i}f_1\sqrt{\text{var}(y_1)} + \alpha_{2i}f_2\sqrt{\text{var}(y_2)} + \dots + \alpha_{pi}f_p\sqrt{\text{var}(y_p)}$$

Las cargas quedan definidas así

$$\lambda_{ij} = \alpha_{ji}\sqrt{\text{var}(y_i)}$$

y los factores específicos

$$\eta_i = \alpha_{m+1i} f_{m+1} \sqrt{\text{var}(y_{m+1})} + \dots + \alpha_{pi} f_p \sqrt{\text{var}(y_p)}$$

Como las componentes principales son no correlacionadas, entonces las η_i no tienen correlación con f_j quedando finalmente

$$x_i = \lambda_{i1} f_1 + \lambda_{i2} f_2 + \dots + \lambda_{im} f_m + \eta_i$$

Sin embargo las η_i si son correlacionadas entre sí.

0.1.2. Factores Principales Iterados

Si la matriz de varianzas y covarianzas corresponde a datos estandarizados ocurre que

$1 = \text{var}(x_i) = \sum_{k=1}^m \lambda_{ik}^2 + \psi_j^2$, de aquí se puede obtener un estimador de la matriz Ψ

Se hace un método iterativo para hallar los factores comunes de la manera siguiente:

1. encontrar una comunalidad inicial.
2. hacer un componentes principales de la matriz $S - \hat{\Psi}$, usar las cargas de los primeros m componentes como columnas de $\hat{\Lambda}$
3. recalcular $\hat{\Psi}$ como la diagonal de $S - \hat{\Lambda}\hat{\Lambda}'$
4. regresar al paso 2 con este nuevo estimador de Ψ

Se itera hasta que las diferencias entre los pares $\hat{\Lambda}$ y Ψ de dos pasos consecutivos sean prácticamente cero.

0.1.3. Factores por máxima verosimilitud

Suponiendo normalidad multivariada en los datos observados y suponiendo el modelo de factores se tiene que la verosimilitud L es función de los datos

y de los parámetros es decir, $L(x; \mu, \Sigma) = L(x; \mu, \Lambda, \Psi)$. En este caso la μ se considera como parámetro de ruido. Y se puede proceder de dos formas, una sustituyendo a μ por el estimador máximo verosímil, y otra descomponiendo la verosimilitud en dos factores, usando verosimilitud condicional dada la \bar{x} . En ambos casos resulta que los estimadores máximo verosímiles $\hat{\Lambda}$ y $\hat{\Psi}$ son los valores que maximizan :

$$F = -\{ \ln|\Lambda\Lambda' + \Psi| + \text{traza}(S|\Lambda\Lambda' + \Psi|^{-1}) - \ln|S| - m \}$$

La función F toma valor de cero si $S = \Lambda\Lambda' + \Psi$ y valores negativos de otra manera. Se usa realmente un método que hace la maximización de la verosimilitud en pasos, ver capítulo 16 de Krzanowski. El método de máxima verosimilitud es el único que es independiente de la escala de medición, da los mismos resultados usando la matriz de correlaciones que la de covarianzas.

0.2. Consideraciones a tomar en cuenta

Los datos están contenidos en la matriz S de tamaño $p \times p$ que contiene $p(p+1)/2$ elementos no redundantes.

Para cada x_i hay que determinar m elementos de Λ_{ij} y un elemento ψ_i

Tambien se introducen $m * (m - 1)$ condiciones para que la matriz $\Lambda'\Psi^{-1}\Lambda$ tenga ceros como elementos fuera de la diagonal.

Entonces para que el modelo sea estimable debe ocurrir que

$p(p+1)/2 \geq pm + p - 1/2 * m * (m - 1)$ En la tabla se muestra que ocurre con distintos valores para p .

Número de variables	3	4	5	6	7	8	9	10	15	20
Máximo número de factores	1	1	2	3	3	4	5	6	10	14

0.2.1. Rotación de factores

Si se tiene una matriz de rotación H esta cumple con $HH' = I$.

Si se rota el factor f como $f^* = Hf$ y se hace $\Lambda^* = \Lambda H'$, resulta que $\Lambda^* f^* = \Lambda H' H f = \Lambda f$, esta f^* cumple con $E(f^*) = H E(f) = 0$ y $Var(f^*) = H Var(f) H' = H I H' = H H' = I$ de manera que el nuevo modelo $x = \Lambda^* f^* + e$ es indistinguible del original $x = \Lambda f + e$.

En muchas ocasiones se usa la rotación para lograr una mejor interpretación de los factores comunes, y sucede que las comunalidades y especificidades se MANTIENEN SIN CAMBIO.

0.2.2. Inconvenientes

1. Hay que decidir cuantos factores hay.
2. Hay limitaciones en cuanto al número de factores según el número de variables originales.
3. Los factores comunes NO son únicos.
4. Si aumenta el tamaño de muestra también aumenta el número de estimadores de factores comunes f_{ij} , m para cada sujeto.