

ANÁLISIS FACTORIAL

Jorge Galbiati R.

El análisis factorial es un modelo matemático que pretende explicar la correlación entre un conjunto grande de variables observadas y un pequeño conjunto de *factores* subyacentes no observadas.

El modelo

Sea X_j ($j = 1, 2, \dots, p$) una serie de variables aleatorias observables. Por ejemplo, la respuesta a la j -ésima pregunta de una encuesta. El modelo supone la existencia de un conjunto f_1, f_2, \dots, f_m de variables aleatorias no observables, tales que:

$$X_j = \mu_j + \lambda_{j1}f_1 + \lambda_{j2}f_2 + \dots + \lambda_{jm}f_m + \eta_j \quad (1)$$

con las siguientes condiciones:

- 1) Los μ_j son las medias de las variables X_j ($j = 1, 2, \dots, p$). Para simplificar la notación se suele estandarizar las variables, de modo que $E(X_j) = \mu_j = 0$. En adelante se trabajará bajo este supuesto.
- 2) Los f_k son independientes e idénticamente distribuidos, y además $E(f_k) = 0$ y $Var(f_k) = 1$ $k = 1, 2, \dots, m$
- 3) Los η_j están independientemente distribuidos. Además $E(\eta_j) = 0$. Supóngase que $Var(\eta_j) = \Psi_j$ $j = 1, 2, \dots, p$
- 4) f_k y η_j son independientes para todo j, k .

Los f_k se denominan *factores comunes*, los η_j *factores específicos*

- 5) Los λ_{jk} son constantes, que se denominan *cargas factoriales*, e indican cuánto pesa cada factor en cada una de las variables.

Se supone, además, que el número de factores, m es mucho menor que el número de variables originales, p .

Forma matricial del modelo

Definimos los siguientes vectores y matrices:

$$\underline{x} = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_p \end{bmatrix} \quad \underline{f} = \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ f_m \end{bmatrix} \quad \underline{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \cdot \\ \cdot \\ \eta_p \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdot & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdot & \lambda_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \lambda_{p1} & \lambda_{p2} & \cdot & \lambda_{pm} \end{bmatrix}$$

El modelo factorial definido en (1) se puede expresar, entonces,

$$\underline{x} = \Lambda \underline{f} + \underline{\eta} \quad (2)$$

y las condiciones 1) a 5) se pueden expresar de la siguiente forma:

- 1) $\underline{f} \sim (\underline{0}, I)$
- 2) $\underline{\eta} \sim (\underline{0}, \Psi)$ en que $\Psi = \text{diag}\{\psi_1, \psi_2, \dots, \psi_p\}$
- 3) \underline{f} y $\underline{\eta}$ son independientes
- 4) $\underline{x} \sim (\underline{0}, \Sigma)$

Se puede ver que $\underline{x} = \Lambda \underline{f} + \underline{\eta}$ implica que $\Sigma = \text{Var}(\underline{x}) = \text{Var}(\Lambda \underline{f} + \underline{\eta}) = \Lambda \text{Var}(\underline{f}) \Lambda' + \text{Var}(\underline{\eta}) = \Lambda I \Lambda' + \Psi$

Entonces queda determinada una forma alternativa de expresar el modelo factorial, que es

$$\Sigma = \Lambda \Lambda' + \Psi \quad (3)$$

De aquí sale lo siguiente: Si $\Sigma = (\delta_{ij})$, las varianzas de las variables originales se pueden expresar como:

$$\text{Var}(X_j) = \delta_{jj} = \sum_{k=1}^m \lambda_{jk}^2 + \psi_j \quad (\text{comunalidad} + \text{especificidad})$$

y las covarianzas entre las variables, como

$$\text{cov}(X_i, X_j) = L_{ij} = \sum_{k=1}^m \lambda_{ik} \lambda_{jk}$$

Para determinar el modelo factorial $\underline{x} = \Lambda \underline{f} + \underline{\eta}$ basta resolver el sistema (3) para Λ y Ψ , dada la matriz Σ .

Este es un sistema no lineal que tiene $p(m+1)$ incógnitas con $\frac{p}{2}(p+1)$ ecuaciones.

Pero de esta forma el modelo está indefinido, pues, tomemos, por ejemplo, cualquier matriz ortogonal $T_{m \times m}$. Por ser ortogonal, es una matriz de rotación. Se tiene que:

$\Sigma = \Lambda \Lambda' + \Psi = (\Lambda T)(\Lambda T)' + \Psi$ y por lo tanto $\Delta^* = \Lambda T$ es otra matriz de cargas que también describe los datos, de la forma

$$\Sigma = \Lambda^* \Lambda^{*'} + \Psi$$

Esto implica que si (Λ, Ψ) es una solución para la ecuación (3), también lo es (Λ^*, Ψ) , con $\Delta^* = \Lambda T$.

Este nuevo modelo equivale a una *rotación* de los factores en el espacio m-dimensional.

Al sustituir Λ por ΛT en la forma original del modelo $\underline{x} = \Lambda \underline{f} + \underline{\eta}$, queda:

$$x = \Lambda \underline{f} + \underline{\eta} = \Lambda T (T' \underline{f}) + \underline{\eta} = \Lambda^* \underline{f}^* + \underline{\eta}$$

$\underline{f}^* = T' \underline{f}$ es una rotación del vector de factores \underline{f} .

Con el objeto de resolver esta indeterminación, se introduce alguna condición inicial arbitraria, como que $\Lambda' D^{-1} \Lambda$ sea diagonal, en que $D = \text{diag} \{ \sigma_{11}, \dots, \sigma_{pp} \}$.

Esto equivale a introducir la condición de que todos los elementos que están fuera de la diagonal de $\Lambda' D^{-1} \Lambda$ (simétrica) son cero. El número de condiciones, entonces, es igual a $\frac{1}{2}m(m-1)$.

Se define el número de *grados de libertad*, s , como el número de ecuaciones más el número de condiciones menos el número de incógnitas. Es fácil ver que

$$s = \frac{1}{2}p(p+1) + \frac{1}{2}m(m-1) - p(m+1) = \frac{1}{2} [(p-1)^2 - (p+m)].$$

El número de grados de libertad s debe ser mayor o igual que cero para que el sistema quede determinado. La tabla siguiente muestra valores de los grados de libertad s , para diversas combinaciones del número de variables originales p y del número de factores m . Las combinaciones que dan valores negativos de s no son factibles. Tampoco se usan aquellas en que el número de factores es mayor que el de variable, pues esto aumentaría la dimensionalidad del problema en lugar de disminuirla.

variables	Número de factores m														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-1	-1	0	2	5	9	14	20	27	35	44	54	65	77	90
2	-1	-2	-2	-1	1	4	8	13	19	26	34	43	53	64	76
3	0	-2	-3	-3	-2	0	3	7	12	18	25	33	42	52	63
4	2	-1	-3	-4	-4	-3	-1	2	6	11	17	24	32	41	51
5	5	1	-2	-4	-5	-5	-4	-2	1	5	10	16	23	31	40
6	9	4	0	-3	-5	-6	-6	-5	-3	0	4	9	15	22	30
7	14	8	3	-1	-4	-6	-7	-7	-6	-4	-1	3	8	14	21
8	20	13	7	2	-2	-5	-7	-8	-8	-7	-5	-2	2	7	13
9	27	19	12	6	1	-3	-6	-8	-9	-9	-8	-6	-3	1	6
10	35	26	18	11	5	0	-4	-7	-9	-10	-10	-9	-7	-4	0
11	44	34	25	17	10	4	-1	-5	-8	-10	-11	-11	-10	-8	-5
12	54	43	33	24	16	9	3	-2	-6	-9	-11	-12	-12	-11	-9
13	65	53	42	32	23	15	8	2	-3	-7	-10	-12	-13	-13	-12
14	77	64	52	41	31	22	14	7	1	-4	-8	-11	-13	-14	-14
15	90	76	63	51	40	30	21	13	6	0	-5	-9	-12	-14	-15
16	104	89	75	62	50	39	29	20	12	5	-1	-6	-10	-13	-15
17	119	103	88	74	61	49	38	28	19	11	4	-2	-7	-11	-14
19	152	134	117	101	86	72	59	47	36	26	17	9	2	-4	-9
20	170	151	133	116	100	85	71	58	46	35	25	16	8	1	-5
21	189	169	150	132	115	99	84	70	57	45	34	24	15	7	0
22	209	188	168	149	131	114	98	83	69	56	44	33	23	14	6
23	230	208	187	167	148	130	113	97	82	68	55	43	32	22	13
24	252	229	207	186	166	147	129	112	96	81	67	54	42	31	21

ANALISIS FACTORIAL MUESTRAL

Si se dispone de una matriz de datos X , constituida por una muestra de tamaño n , de observaciones de p variables, se calcula la matriz de varianzas-covarianzas muestral

$$S = X'(I - \frac{1}{n}1_n1_n')X$$

y se resuelve el sistema (no lineal)

$$S = \Lambda\Lambda' + \Psi \quad \text{para } \Lambda \text{ y } \Psi$$

Con con la restricción inicial de que $\Lambda'D^{-1}\Lambda$ sea diagonal, con $D = \text{diag}\{s_{11}, \dots, s_{pp}\}$, la parte de la diagonal de S .

Hay varias maneras de hacerlo, siendo los métodos más conocidos el del *Factor Principal* y el *Máximo Verosímil*.

ESTIMACION POR EL METODO DEL FACTOR PRINCIPAL

Se asumirá que las variables están estandarizadas, lo que corresponde a usar la matriz de correlaciones R en lugar de la matriz de varianzas-covarianzas S .

Se inicia haciendo estimaciones iniciales de las especificidades ψ_j . Hay más de una forma de hacerlo, siendo una de ellas el mayor coeficiente de correlación entre x_j y las demás variables.

Sean $\tilde{\psi}_j$ estas estimaciones iniciales. Por el teorema de descomposición espectral, se puede expresar

$$R - \tilde{\Psi} = \sum_{i=1}^p \lambda_i \underline{\gamma}_i \underline{\gamma}_i'$$

con $\tilde{\Psi} = \text{diag} \{ \tilde{\psi}_1, \tilde{\psi}_2, \dots, \tilde{\psi}_p \}$ y con $\lambda_1 > \lambda_2 > \dots > \lambda_p$ los valores propios y $\underline{\gamma}_i$ los respectivos vectores propios de $R - \tilde{\Psi}$

Si los primeros m valores propios son positivos, se tiene que

$$\hat{\lambda}_{(i)} = \lambda_i^{1/2} \underline{\gamma}_i$$

En forma matricial

$$\hat{\Lambda} = T_1 \Lambda_1$$

en que T_1 es la matriz cuyas columnas son los vectores propios ordenados,

$$T_1 = (\underline{\gamma}_1, \underline{\gamma}_2, \dots, \underline{\gamma}_p), \text{ y } \Lambda_1 = \text{diag} \{ \lambda_1^{1/2}, \lambda_1^{1/2}, \dots, \lambda_m^{1/2} \}$$

A partir de aquí se puede obtener estimaciones revisadas de los ψ_j a partir de los elementos de la diagonal de $R - \tilde{\Psi}$, mediante

$$\tilde{\psi}_j = 1 - \sum_{i=1}^m \lambda_{ij}^2$$

Se repite el procedimiento y se sigue iterando hasta lograr convergencia. Si algunos de los primeros m valores propios son menores que cero, es una indicación de que hay exceso de factores en el modelo, y hay que reducir m .

ESTIMACION POR MAXIMA VEROSIMILITUD

Hay que asumir una familia de distribuciones. El supuesto utilizado es de normalidad para las filas de la matriz de datos X . Entonces $\underline{x} \sim N_p(0, \Lambda\Lambda' + \Psi)$ representa la población de donde se obtuvieron las observaciones.

La log-verosimilitud se puede representar como

$$l = -\frac{1}{2}n \log |2\pi\Sigma| - \frac{1}{2}n \text{traza}(\Sigma^{-1}S)$$

en que $\Sigma = \Lambda\Lambda' + \Psi$. La log-verosimilitud se debe maximizar con respecto de Λ y Ψ para obtener los estimadores.

Maximizar la verosimilitud es equivalente a minimizar la función

$$F(\Lambda, \Psi) = p(a - \log g - 1)$$

en que a es la media aritmética y g la media geométrica de los valores propios de $\Sigma^{-1}S$.

La minimización de F entrega los estimadores máximo verosímiles de Λ y Ψ . Esto se hace minimizando F con respecto de Λ , para Ψ fijo y luego minimizando con respecto de Ψ . La primera minimización se hace analíticamente, la segunda debe hacerse por métodos numéricos.

Los resultados obtenidos por los dos métodos, *factor principal* y *máxima verosimilitud* no son iguales.

PRUEBA DE HIPOTESIS DE BONDAD DE AJUSTE

Si se asume normalidad para las filas de X y se obtienen los estimadores máximo verosímiles, se pueden usar para probar la hipótesis:

$H_0 : \Sigma$ arbitraria sin restricciones, versus

$H_1 : \Sigma = \Lambda\Lambda' + \Psi$

Para m factores, el estadístico de prueba es

$$nF = np(\hat{a} - \log \hat{g} - 1)$$

con \hat{a} y \hat{g} medias aritmética y geométrica de los valores propios estimados de $\Sigma^{-1}S$.

nF tiene distribución asintótica χ^2 con $s = \frac{1}{2}(p-m)^2 - \frac{1}{2}(p+m)$ grados de libertad, cuando H_0 es verdadera.

ROTACION DE FACTORES

Como se dijo antes, si Λ es una matriz de cargas factoriales, ΛT también lo es, si T es una matriz ortogonal, lo que equivale geoméricamente a *rotar* los factores.

En efecto, si el modelo factorial es $\underline{x} = \Lambda \underline{f} + \underline{\eta}$, resulta equivalente al que resulta de sustituir Λ por ΛT , quedando como $x = \Lambda T(T'f) + \eta$

$T'f$ corresponde a una rotación del vector de factores f en el espacio m -dimensional.

Este hecho se aprovecha para seleccionar un conjunto de factores Tf optimal desde el punto de vista de su interpretación.

El método de rotación más popular es el llamado VARIMAX, que busca que la matriz de cargas resultante tenga columnas con muchos ceros, para facilitar la interpretación de los factores. De este modo, en la expresión (1), aparecerán sólo algunos f_k significativos, los demás serán despreciables.

Si $B = \Lambda T$ es la matriz de cargas rotada, $B = (b_{jk})_{p \times m}$, el criterio Varimax busca maximizar la suma de las varianzas de los b_{jk}^2 dentro de cada columna.

La cantidad a maximizar es

$$V = \sum_{k=1}^m \sum_{j=1}^p \left(b_{jk}^2 - \frac{1}{p} \left(\sum_{j=1}^p b_{jk}^2 \right) \right)^2$$

$$V = \sum_{k=1}^m \sum_{j=1}^p b_{jk}^4 - p \sum_{k=1}^m \bar{b}_k^2$$

en que \bar{b}_k^2 es el promedio de los b_{jk}^2 de la columna k

$$\bar{b}_k^2 = \frac{1}{p} \sum_{j=1}^p b_{jk}^2$$

Esto hace que los b_{jk}^2 sean lo más discímiles posible. Como no pueden ser menores que cero, resultarán algunos b_{jk} cercanos a cero y otros grandes (en valor absoluto).

Nótese que

$$\Lambda \Lambda' = (\Lambda T) (\Lambda T)' = B B'$$

por lo tanto la comunalidad (la parte de la varianza que corresponde a los factores comunes) se mantiene invariante:

$$h^2 = \sum_{k=1}^m \lambda_{jk}^2 = \sum_{k=1}^m b_{jk}^2$$

Otros métodos de rotación que buscan resultados similares son:

QUARTIMIN
 QUARTIMAX
 TRANSVARIMAX
 EQUAMAX
 RATIOMAX
 PARSIMAX

Los que parten de criterios distintos, pero llegan a resultados similares, aunque no iguales.

SCORES FACTORIALES

El problema siguiente es, dados los valores particulares x_j ($j = 1, 2, \dots, p$) de las variable correspondientes a un caso (por ejemplo, las respuestas de un individuo a cada una de las preguntas de una encuesta), cuánto son los valores de los factores comunes f_1, f_2, \dots, f_m y del factor específico η_j en ese caso, para que se cumpla que

$$x_j = \lambda_{j1}f_1 + \lambda_{j2}f_2 + \dots + \lambda_{jm}f_m + \eta_j$$

Una forma de obtenerlos es asumir que no son aleatorios, sino parámetros que se pueden estimar. Se parte del vector de derivadas de la log-verosimilitud igualado a cero, que se usó para estimar los parámetros del modelo,

$$\frac{\partial l}{\partial \underline{f}} = \Lambda' \Psi^{-1} (\underline{x} - \Lambda \underline{f}) = \underline{0}$$

de donde

$$\underline{f} = (\Lambda' \Psi^{-1} \Lambda)^{-1} \Lambda' \Psi^{-1} \underline{x}$$

Estas estimaciones se denominan *Scores Factoriales de Bartlett*.

También están las estimaciones bayesianas, dada la distribución a priori $N_m(\underline{0}, I)$ para el vector de factores \underline{f} , que resultan ser

$$\underline{f} = (I + \Lambda' \Psi^{-1} \Lambda)^{-1} \Lambda' \Psi^{-1} \underline{x}$$

y dan valores ligeramente distintos que las de Bartlett.

Finalmente, los scores de η_j se obtienen por diferencia,

$$\eta_j = x_j - \lambda_{j1}f_1 + \lambda_{j2}f_2 + \dots + \lambda_{jm}f_m$$